Aesthetic Information Collages: Generating Decorative Displays that Contain Information

James Fogarty[†], Jodi Forlizzi[†*], and Scott E. Hudson[†]
[†]HCI Institute, and [*]School of Design, Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213
{jfogarty, forlizzi, scott.hudson}@cs.cmu.edu

ABSTRACT
Normally, the primary purpose of an information display is to convey information. If information displays can be aesthetically interesting, that might be an added bonus. This paper considers an experiment in reversing this imperative. It describes the *Kandinsky* system which is designed to create displays which are first aesthetically interesting, and then as an added bonus, able to convey information. The Kandinsky system works on the basis of aesthetic properties specified by an artist (in a visual form). It then explores a space of collages composed from information bearing images, using an optimization technique to find compositions which best maintain the properties of the artist's aesthetic expression.

Keywords
Visual design, aesthetics in computational objects, display generation, ambient information displays in decorative objects, optimization, simulated annealing.

"…But does it go with the couch?"

INTRODUCTION AND MOTIVATION

As computer use has shifted into wider aspects of life, the requirements that it has faced have shifted as well. The value of computing technology was traditionally measured by its results – largely its *usefulness* in solving problems of interest. With the advent of personal computers came the important imperative of *usability*. As computational technology moves beyond the confines of the work environment and into the rest of our lives, we have begun to see an additional requirement emerge: *desirability*. Products such as the Apple iMac have shown that selling computers is starting to be about "cool" and "interesting" and even "beautiful", as well as "understandable", "easy to use" and "powerful".

As computing continues its transformation from specialized tool into everyday object, we expect the role of desirability to increase. For example, if we were to buy a low-end computer to place in our living room today, it is fairly likely that it could be purchased for less money than was spent on the furniture nearby (and it certainly will need to be replaced sooner). In this environment, new questions start to arise such as: "*Does it go with the couch?*" As yet, we have only rarely addressed questions of this sort – which have issues of visual design and aesthetics at their core – as a central part of interactive system designs. The work described here considers technology that may be useful when approaching systems from this viewpoint.

This paper considers the *Kandinsky* system for generating *aesthetic information collages*. An aesthetic information collage is designed to be a type of ambient information display in a decorative object (see also [5]). Normally, the central imperative for an information display is to effectively convey information – if it can also be aesthetically interesting, that might be an added bonus. The Kandinsky system is an experiment in turning this imperative upside-down. We envision this system being used in a home or office setting to produce the equivalent of a painting or poster hanging on the wall. (With current advances such as low cost Organic LED displays, this may be widely practical in only a few years.) Like other images we hang on the wall, we would typically choose this display primarily because of its aesthetic properties. To these aesthetic properties, we wish to add the ability to convey ambient information – information that we may wish to be aware of, but that should not necessarily demand our attention.

Kandinsky works by composing images representing information items to be displayed into a collage in a way that tries to maintain certain aesthetic properties (for other approaches to automatic generation of collages, see for example [4, 9, 14, 20]). If necessary, the system will also create images to represent information given to it in only textual form.

Figure 1 shows an example of a collage created by the Kandinsky system. This collage represents a message from the comp.human-factors news group, with the title: "CfP Ubicomp 2001 - Late breaking Results, Workshop …". It was generated by first creating a set of images based on the textual message subject string using the *ImageConjure* subsystem. These images were then combined using an *aesthetic template* to form the final collage. We envision that a series of these collages would be generated from the

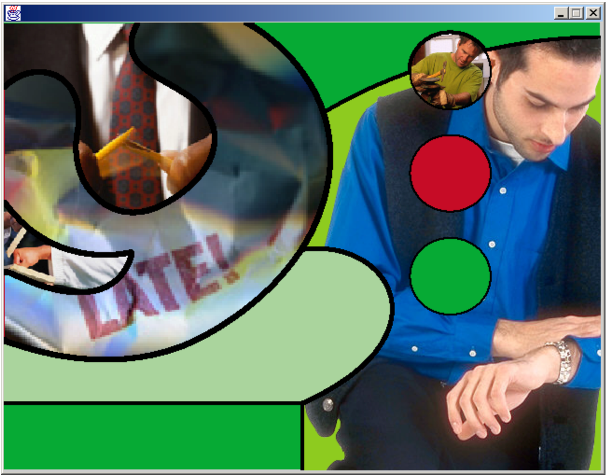

Figure 1. Example Collage Generated From:
"CfP Ubicomp 2001 - Late breaking Results, Workshop…"
(Please see the ACM digital library for color versions of all figures.)

same template over time, each representing a small number of messages. We would then slowly shift between the collages to present a larger body of information.

Artists exhibit a capacity for creativity and aesthetic judgment that is poorly understood and not currently well suited to automation. Rather than attempt to build robust knowledge of aesthetic principles into the system, or to create a form of artificial artist (a very challenging task, see for example [7, 1]), we have chosen instead an approach that allows the artist to express aesthetic concepts in visual form – an *aesthetic template*. Aesthetic templates allow an artist to express overall compositional form, dominant colors, and other aesthetic properties that the final collage is to follow, along with fixed elements of its content. We then attempt to reuse that expression by casting the collage generation process into an optimization problem – one of matching and manipulation of the visual properties of potential collage elements in order to mimic the properties of various parts of the template. In this way, we use the computer for a repetitive algorithmic task that it is well suited for in order to leverage and amplify the knowledge and skill of a human artist.

In the next section we will present an overview of the components and basic action of the Kandisnsy system. We will then consider a simple technique for fabricating images that are likely to be representative of information given in only textual form. This is followed by a discussion of the extensible optimization process that forms the core of the Kandinsky system. Next, a discussion of visual properties will be provided, followed by a detailed discussion of the aspects actually used by the system, and how they are implemented. We will then show some results of the system and provide a conclusion.

ARCHITECTURAL OVERVIEW

Figure 2 illustrates the architecture of the Kandinsky system. Typical use begins with the arrival of information that we might want to be displayed ambiently. In the work described here, we have chosen electronic messages (e.g., email or news group postings) as a ready and familiar source of information, but many other sources could be used instead. From the incoming information, we extract textual summaries such as message subject lines, as well as any suitable images that might come with the information (for the examples in this paper we have chosen to work only from text to illustrate the more challenging problem). When images are not provided, we use textual information with the ImageConjure subsystem (described in the next section) to find images within a large indexed image collection that are likely to reflect the semantic content of the message. The Kandinsky system then composes these images on the basis of an aesthetic template to create a collage.

As discussed in the section after next, an aesthetic template is defined using a collection of layered *regions* supplied by an artist. Regions can directly contribute image components to the final result, can specify layout and matching criteria, and/or can specify image manipulations to be performed.

The Kandinsky system uses the layout and matching criteria of the template (for example, color or texture matching against the template image) as the definition of an optimization problem. In particular, each criteria in the system comes with a scoring function which evaluates how well the criteria holds within a given composition or partial composition. Roughly speaking, the system generates a series of candidate assignments of images to regions and a layout within those regions. Based on the matching and

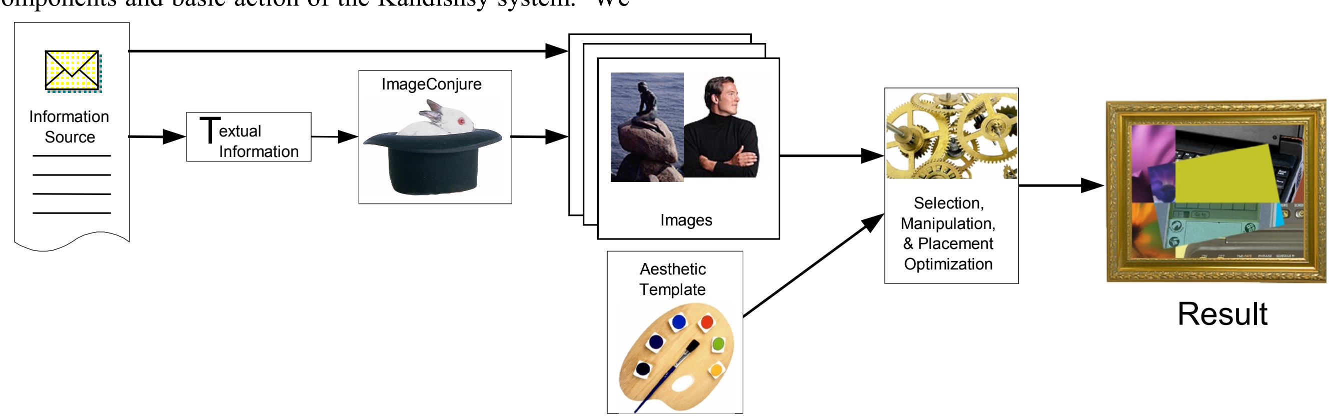

Figure 2. Architecture of the Kandinsky System

layout criteria specified in the template, it does a simulated annealing optimization to improve the goodness of the match and layout. This optimization process is iterated until a fixed time limit is reached. At this point any image manipulations (such as clipping to a region mask, applying a texture, or modifying colors) that are specified by a region are performed. Finally, the regions are rendered as a series of layers in order to produce the final result.

By casting the composition algorithm as a general optimization problem, we have created a system that is extensible and flexible. With suitable matching metrics, it can handle a number of the properties of importance in visual design. In addition, new properties can be added when new metrics are invented, without disturbing existing templates or regions. Further, while in the examples presented here we have concentrated on visual properties, other kinds of metrics can be applied as well. For example, regions might include criteria based on the type or content of information being represented. This would allow, for example, the images associated with email messages from senders on a list of close colleagues to be placed in a particular location simply by creating a region which "has an affinity" for that kind of information (as expressed by its optimization criteria). This allows both aesthetic and communicative properties to be expressed in the same framework, and allows different templates to trade off these criteria in different ways. Finally while seemingly complex and technical, this framework as we will see later in the paper, in fact lends itself to a simple artist's interface which hides many of the details behind a metaphor of paint and painting.

REPRESENTATIVE IMAGES

The Kandinsky system works by composing images. However, many of the information sources we might wish to display ambiently – such as the state of our unread email – are primarily textual in nature. The ImageConjure subsystem is responsible for overcoming this limitation by converting text into a set of representative images.

The authors first came across this *image summarization* problem in our experiences using the *What's Happening* community awareness system [21], where some information sources (e.g., Slashdot article excerpts [17]) had iconic subject indicators with them because a human editor had chosen an appropriate image as a part of the publishing process. We found these to be very helpful in assessing the information at a glance. However, for the majority of the information presented, the luxury of a human editor was not available.

The job of the ImageConjure subsystem can be seen as providing some approximate replacement for a human editor picking summary images. While this at first blush seems like an exceedingly hard task, we realized that the human knowledge necessary to do this had, in some sense, already been captured. In particular, there are large collections of stock photo and clip art images available that have been topically indexed. The ImageConjure subsystem makes use of this prior human effort via some simple information retrieval technology. It uses a textual summary of the information to be represented (e.g., the subject of a message – which has also been prepared by a person) as a search string against the indexing information provided with several large image collections.

For our current implementation, we have acquired a license for use of two indexed image collections. The first collection (from PhotoDisc Inc. [13]) contains approximately 24,000 royalty-free stock photos. This collection is at relatively low resolution (primarily 278x183 pixels). However, this is generally suitable for our purposes and was available at reasonable cost. This collection has the significant advantage that nearly all photos have been carefully composed by professional photographers. As a result, they have relatively high production values and generally contribute positively to our overall result.

This collection, however, has at least one drawback. In some cases, it seems to be "over indexed" for our application – in particular, peripheral or background objects appearing in the photos are included in the index. This causes images to be selected that contain an appropriate object in the background, but whose dominant object or theme is unrelated to the query at hand. This is an instance of a slightly more general problem. It stems from the fact that information retrieval systems of all sorts have been tuned primarily to uses where humans make a final judgment. As a result, they typically tend to reduce the precision of their results in various ways in order to attain greater recall.

Our second collection (licensed from Hemera Inc. [6]) contains approximately 50,000 images of background-separated objects (i.e., each image comes with a mask separating its object(s), from the background). This collection has the advantage of non-rectangular images and, because of its primarily single-object images, queries against it seem to typically have higher precision in their topical match. However, the images in this collection have somewhat lower production values and do not typically represent an interesting composition in and of themselves. Also we found that in this very large collection there were sometimes several different photographs of the same object (e.g., taken from different points of view). Because of the matching process, if one of these images scored highly, the other variations tended to be returned as well. In some cases this squeezed out images of other objects which would have been more interesting than a repeated object.

Based on some experimentation with the properties of these two image collections, we have tentatively settled on giving a significantly higher weight to the selection of images from the Photodisc collection – primarily because of the typically higher aesthetic value of the pictures it contains. Specifically, we have been using a strategy of selecting one image from the Hemera collection and three or four images from the Photodisc collection.

Searching by the ImageConjure system is done using the Lucene indexing and search package (available free with Java source code on the web [2]). However, many information retrieval systems would probably produce similar results, and more sophisticated systems might offer somewhat improved outcomes.

Overall, we found the results of the ImageConjure to be quite useful, although not perfect. Figure 3 gives a representative sample of results. The queries used here are taken from the subject lines of messages found in the comp.human-factors news group (these titles were simply the first ones on the list the day we collected them). Within these examples, we can see many relevant images. For example, the selection of images related to Denmark and the pictographic equivalent of "late breaking" both work very well. In addition, we can see several typical (partial) failures. For example, the results for "Seeking Entry-level HCI work" show a typical issue with alternate meanings (here we have an entry-way and a level for "entry-level", and slightly lower ranked images also return religious themes based on "seeking"). This instance also illustrates the repeated object problem.

In addition to the 10 queries shown below, two other queries from our first 12 failed to retrieve any images from our collections (these were: "any interesting GUI interfaces?" and "TM"). In cases where the message subject fails to return a result, we can consider several different fallbacks. First, we search based on words in the message body. In these cases, both messages returned multiple images based on the first 10 words of the message. If this fails, we currently use a "not found" image randomly selected from a small collection of appropriate images (which we formed based on results from the query "question missing gone mystery"). We could also employ a

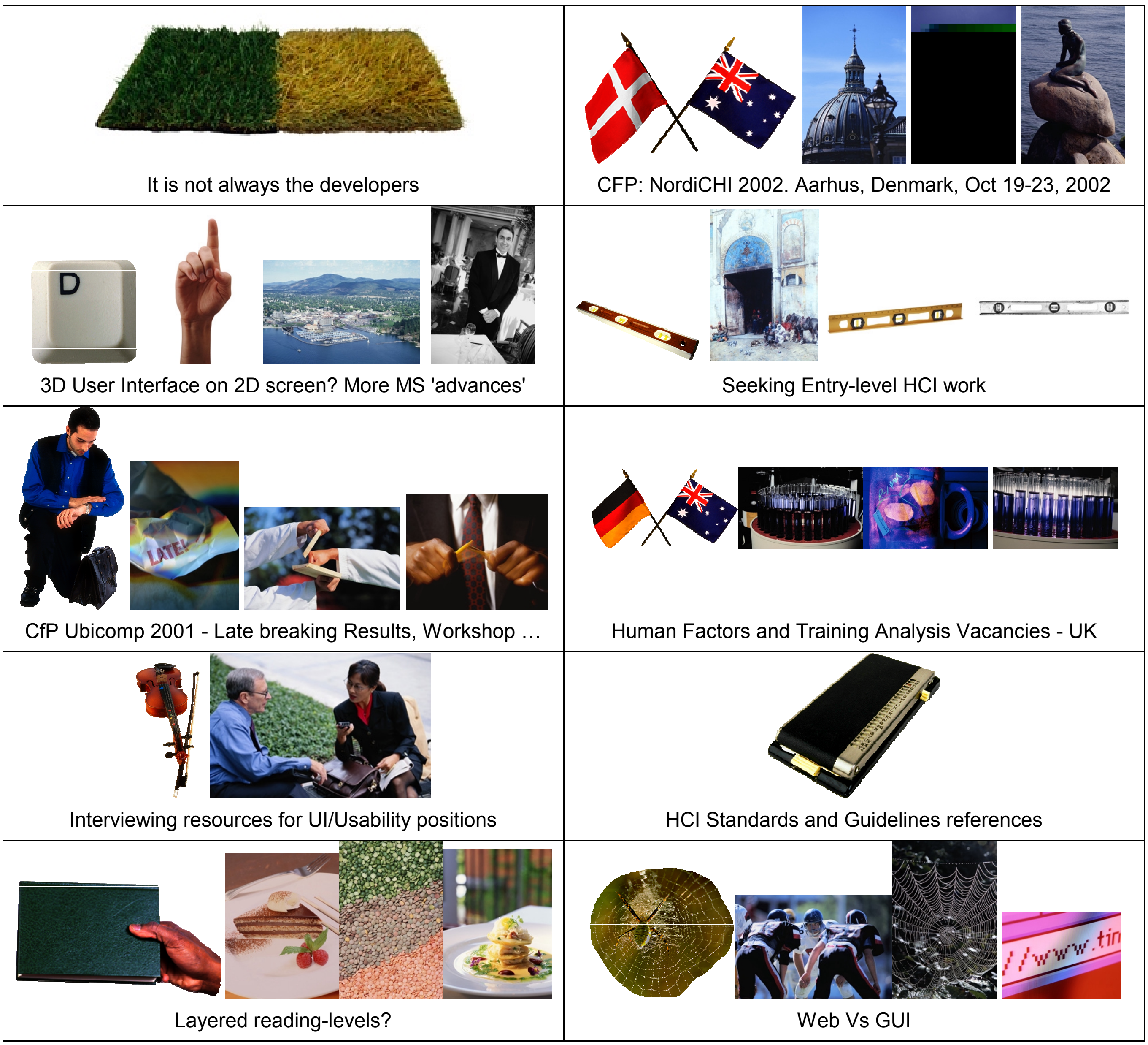

Figure 3. Representative Results from the ImageConjure Subsystem.

fallback strategy of rendering the text itself to form an image in various interesting ways.

THE OPTIMIZATION PROCESS

The optimization mechanism that forms the core of the Kandinsky system takes a general simulated annealing approach. This approach works with a configuration of components (comprised of properties such as the selection of information images, their placement within the collage, etc.) Over the course of the optimization, it attempts to improve that configuration by iteratively perturbing the configuration in some way, then evaluating the result. In normal simulated annealing fashion, a "temperature" parameter is maintained which controls the perturbation and iterative direction of the system. The temperature is initially set high, indicating that large perturbations are to be employed, and that the system may retain some perturbation steps that actually decreased the assessment of the overall system. Over time, the temperature is decreased. This indicates that smaller perturbation steps are to be taken, and that only steps which improve the assessment of the overall system will be retained.

Aesthetic templates are the central mechanism for expressing the aesthetic properties which the simulated annealing optimization operates over. An aesthetic template is composed from a layered set of regions. As detailed below, regions may specify four things: Fixed visual elements, an initial selection and placement strategy, a series of evaluation criteria, and a set of post-processing effects.

Fixed visual elements: These provide specific image content that is added directly to the collage, and forms a backdrop to any information images associated with the region. In some cases, artist provided visual elements actually dominate the composition. However, the degree to which the final collage is the result of direct composition by the artist, versus composed from information images, is something fully at the control of the artist. Further some regions will be solely for composition or layout purposes, and will provide no visual elements of their own.

Initial image selection and placement strategy: This is a mechanism for initially selecting images which are to be associated with a region, and initially placing them (*laying them out*) within the region. It provides the details necessary for the initialization of the optimization process at a particular configuration. The basis for selections might be visual in nature, or might be semantic. Similarly a variety of initial placement strategies might be employed. (At present the system only supports one image selection and placement strategy: a random one, relying instead on the evaluation and optimization process alone to produced the desired results. However, this can be inefficient, and more targeted strategies will be added as the system matures.)

Evaluation criteria: Central to each region is a set of evaluation criteria. These criteria measure the goodness of a candidate configuration for the region. The artist indicates which evaluation criteria will be applied for each region in order to describe which visual properties are desired for the region. Typical criteria include layout criteria (e.g., favoring non-overlapping layouts, layouts staying within a region mask, or layouts conforming to a grid structure), as well as visual matching criteria (e.g., criteria for matching color, texture, dominant image direction, or balance requirements). With each evaluation criteria comes a system-supplied scoring function, and an artist-supplied weight expressing how important the criteria is. Evaluation criteria can be provided in three forms. The criteria can be applied on a per image basis (evaluating aspects of the inclusion, placement or other properties of a single image in the region), or on a per region basis (evaluating the goodness of the entire region configuration). In addition, evaluation criteria that apply to the entire collage can be provided as part of the aesthetic template.

The final composition process for a collage involves placing the selected information images for a region in their optimized configuration (e.g. at their optimized positions) over the fixed visual elements of the region. Post-processing effects are then applied to the resulting composition.

Post-processing effects: Each region may provide one or more post processing effects that are to be applied after its final configuration is determined. Typical effects include clipping to a mask supplied with the region, and applying color and texture changes. The results from each region are then rendered as an ordered set of layers to create the final collage.

Optimization proceeds by selecting and applying a perturbation operator. The current temperature is typically used by these operators to determine how large a perturbation to generate, and to balance between gradient-decent steps (at lower temperatures) and random steps (at higher temperatures). The resulting configuration is then scored using the local, regional, and global evaluation criteria and associated weights from the aesthetic template. Generally, if a perturbation improves a configuration, the new configuration is retained for future steps, otherwise it is discarded (this corresponds to a simple hill-climbing optimization strategy). However, as is characteristic of simulated annealing optimizations, in order to avoid being trapped at less desirable local maxima, the system will also allow the retention of "backward" steps with a small probability which is proportional to the current temperature.

The Kandinsky system provides an extensible library of mechanisms, evaluation criteria, and post-processing effects that can be made available to the artist for inclusion in an aesthetic template. In addition, the system contains an extensible library of perturbation methods each of which modifies some aspect of a candidate configuration. For example, perturbation operators are provided for reassigning images to different regions, and for moving the spatial position of images within regions.

The next section describes properties from principles of visual design that we might use to create such a library in a principled way. It also considers how some of these properties might be operationalized, by either measuring them (for use in evaluation criteria to drive optimization), or being able to generate images which embody them (for use in perturbation or post-processing methods).

PROPERTIES OF INTEREST

Since the earliest civilization, humans have communicated and solved problems through the use of visual information. Visual composition is critical in both communicating and problem-solving. However, it is not well known how artists and designers create visual compositions, nor is it easy to replicate. Artists and designers often "operate from the gut" and are not able to fully articulate what compositional decisions they have made and the rationale for doing so.

However, much as syntax has evolved as a way to bring order and structure to language, researchers in art and visual perception have evolved a syntax of visual literacy – a set of concepts and principles for understanding and creating structure in visual compositions (see for example [3, 10, 19]). Although the details of these constructs vary somewhat in form and emphasis, they have a good deal in common because they are all derived from the same underlying human sensory, perceptual, and cognitive phenomena. For the work presented here, we have adapted (paraphrasing, slightly reorganizing, and simplifying) the constructs presented by Dondis [3] as a guide for understanding what aspects of visual composition we should consider addressing.

Following Dondis, there are two levels at which important properties for visual compositions emerge. One level considers primarily low-level image features and corresponds roughly to a sensory level. The second level concerns itself primarily with the psychological reactions that are induced by images. This corresponds roughly to perceptual and cognitive levels of human image perception and understanding.

At the lower level are five primary properties of interest. These include: color, texture, "edges and lines", direction, and shape. Five additional important properties emerge at the higher level. These include: "relative contrast", dimensionality, balance, motion, and stress. Each of these properties is briefly considered below and is related to potential computational treatments of them.

Color is an invaluable source of information in perception of the environment. Color is also often used to create symbolic or emotional meaning. For example, red can represent blood. Therefore, we see color as information operating on universal, cultural and even personal levels. Color is a familiar concept computationally, and color matching and manipulation primitives are readily available for use.

Texture is the visual element that best represents what we might feel if we were to touch a surface. Texture in a visual composition takes the form of minute variations on a surface or artifact. Our perceptual systems are extremely sensitive to detecting different textures. For example, false texture is used by animals in nature to confuse possible predators. Significant work in computer vision has been done on analysis and synthesis of textures (see [18] for a survey) and a number of these techniques can be adapted for use in our context.

Edges and Lines are fundamental to our perception of boundaries, and hence to the perception of separate objects. As a result, lines – both in isolation and as formed by the juxtaposition of contrasting image components – represent strong and important visual components. Again, because they are fundamental perceptual constructs, computer vision research provides a ready set of techniques for dealing with lines and edges in images.

Direction. Moving to a slightly higher level, lines and edges induce a sense of directionality within an image. The horizontal and vertical (which are reinforced and given special importance by gravity and the horizon in the physical world) form an important frame of reference for directionality in visual composition. Again, simple computer vision techniques for extracting approximations of directionality information are readily available.

Shape is induced by lines or edges in closed form. Visual designs are often expressed in terms of the composition of basic shapes such as squares or rectangles, circles or ovals, and triangles. For example, when learning to draw the human figure, teachers will abstract the form into a series of ovals and rectangles. Shapes can also impart meaning. For example: "The square has associated [with] it dullness, honesty, straightness, and workmanlike meaning; the triangle, action, conflict, tension; the circle, endlessness, warmth, protection." [3, p. 44].

Computationally, measurement of shape has at its core a notion of separation and segregation of objects. Unfortunately, tasks such as robust separation of foreground objects from a naturalistic background (figure-ground separation) are beyond the current abilities of computer vision systems in most cases. However, on the generative side, we can take advantage of readily available capabilities for masking or clipping to create shapes of interest. In addition, it may be possible to use shape-oriented metrics such as (approximate) fractal dimension [11] without explicit object separation.

Relative Contrast concerns the relationship of objects with respect to some property with magnitude (such as size or brightness). This property is evident in the forms of *proportion* and *scale*, which are contrasts of size between different objects, and between objects and the overall visual field, respectively. These properties can have an important impact on the effects of a composition. For example, a small image of a person surrounded by a vast landscape evokes a different feeling than a portrait at close range.

Computational measurements of contrasts in color or brightness are relatively easy to perform. On the other hand, computational measurement of scale and proportion are again impeded by a lack of robust object separation techniques. However, on the generative side, manipulation of size properties via image scaling is well understood.

Dimensionality. The human perceptual system does a remarkable job of perceiving three-dimensional structure from two-dimensional images. Dimensionality refers to the extent to which an image induces a sense of depth, and to the mechanisms (such as different forms of perspective) by which this is accomplished. At present, little is available computationally to help with notions of dimensionality as they arise in this context.

Balance. At a higher level, we intrinsically estimate properties of physical objects from their images. One such property is balance. Man's need for balance is an important physiological as well as physical influence. As a result, equilibrium, in the form of balance, is a strong visual referent – we rely on horizontal and vertical axis when interpreting a visual composition, and feel comfortable if those axes are reinforced in the composition via balance with respect to them. A variety of moment and center of gravity calculations can be used to estimate balance properties of images.

Motion. At a similarly high level, we intrinsically estimate properties of implied movement from aspects of still images. These properties take their imperative from the physics of the world and (implied) movement brings with it meanings from the physical world. Computationally, we have few tools for measurement of implied movement within still images, but generative capabilities such as simulation of motion blur are available.

Stress (vs. Stability). At an even higher level of abstraction, properties such as balance, motion, shape, and contrast, can induce feelings of stability (and when lacking, tension or stress) in a composition. For example, compositions which lack balance, or which shift the implied axis of balance or symmetry far away from the center, induce a feeling of tension or stress which conveys subtle meaning. Since stress as described above is not a well understood concept, it is not surprising that few computational approaches for measuring or generating it are readily apparent.

As a general trend, we can see that at the lower levels of abstraction, the principles we have outlined here can be more easily operationalized in either an evaluative form (that can be used to drive optimization) or a generative form (that can be used in perturbation, or more commonly post-processing effects). At the higher levels of abstraction we may only have available techniques that can generate some limited forms of various properties.

EVALUATION AND GENERATION SPECIFICS

In this section, we describe the specifics of how aesthetic properties are evaluated and generated within the system. Recall that the system deals with aesthetic properties in four ways: initial image selection and placement methods (used to initialize configurations prior to optimization), evaluation criteria and perturbation rules (used to drive improvement of aesthetic property matches), and after-effects (used to generate desired properties).

As indicated earlier, in our initial implementation of the system, only simple random initial selection and placement methods are provided – we instead rely on subsequent perturbation steps to achieve desired properties. We also presently support only a minimal set of perturbation methods – a method for switching the assignment of images to layers, and another for image placement within regions. Since the system is fully pluggable, these limited methods can easily be extended with more efficient and sophisticated methods in future versions.

The library of evaluation criteria offered by the current system is more extensive. Criteria are provided that operate at the level of individual images (e.g., scoring the placement of a particular image at a particular point within a region), at the level of regions (i.e., scoring the goodness of the current configuration of a region), and at the level of the whole composition (i.e., scoring some property that is measured across the configuration of the entire collage).

At the individual image level, criteria are provide for: containment of the image within a region's clipping mask, aspect ratio match (compared to the target aspect ratio of the region), local color match (compared to the region image in the neighborhood of a particular placement location), region color match (compared to a region target color), direction match, and texture match.

Figure 4 shows an example of color matching along with balance, "single image", and "stay within mask" criteria. Here each bottle is a separate region employing color matching against the dominant color of the region image, as well as expressing an affinity for exactly one image which is contained within its clipping mask. Each region also specifies a balance point, resulting in the 'V' placement.

Direction match is performed by doing edge extraction on the region image with a local gradient operator, then classifying local direction of the edges as horizontal, vertical, left diagonal, right diagonal, or none. The distribution of these edge classifications is then compared

Figure 4. Example of Color Matching and Balance

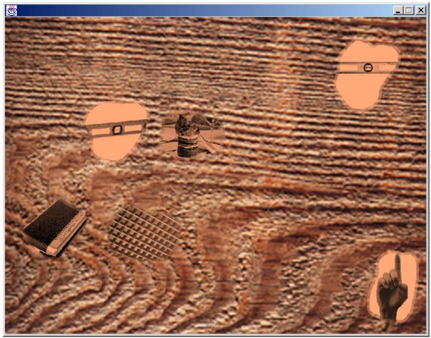

Figure 5. Example of Directional Matching

with a similar distribution computed for candidate images to be placed over the region. Figure 5 shows an example of direction matching where the background wood grain provides the direction matched against.

Texture matching is done via the Local Binary Pattern (LBP) and Local Binary Pattern Contrast (LBP/C) classification methods described and evaluated in [12].

Evaluation criteria currently operating at the region level include: number of items placed in the region (comparing to preferred upper and lower limits), non-overlap of items, color match (compared to overall color of the region), and balance.

Balance is computed by calculating the center of mass with respect to an edge density metric. We also intend to experiment with an image segmentation-based center of mass metric. Centers of mass calculated for images and regions are also used to estimate image centers and are employed in several after-effects described below. Figure 4 provides an example of the use of balance criteria. Here the five regions are all center balanced horizontally and set to top, center, bottom, center, and top-heavy respectively.

Evaluation criteria operating at the level of the full composition include: non-overlap of images, and balance.

After-effects currently supported include: colorization (reducing an image to grayscale, then tinting it with color(s) coming from the region), clipping of images to their region, scaling, and blur. Scaling is done around the calculated center of an image (typically the center of mass as calculated for balance). The overall set of after-effects that might be applied is quire large due to the proliferation of commercial effect filters (e.g., Photoshop plug-ins). Nearly any of these effects might be considered for the after-effect library.

Figure 6 provides a complete example combining several evaluation criteria and after-effects. This template is based on the Kandinsky painting "Composition IV" [8]. As illustrated in the lower center, several regions supply fixed images from the original painting, while the remainder of the regions in the template are set for color matching, and direction matching (with equal weights in one or two directions). In addition, each of these regions applies an aspect ratio match. Finally, these regions also apply an after-effect of colorization using colors from the original painting as illustrated in the upper center. The final result when combined with images (left) representing "Web vs. GUI" and "BayCHI 5/8/01 Using Ethnography in the New Economy", is shown at the right.

The set of methods, criteria, and effects described above provide reasonably good coverage of the basic visual properties considered in the previous section. However, a key feature of the Kandinsky system is its extensibility. This will allow us to extend and enhance the current library in an incremental fashion without changing the underlying system architecture.

In addition, to the basic visual properties targeted above, we have also experimented with an evaluation criteria based on recognition of a very high-level visual construct which has significant aesthetic impact: *the human face*. For this evaluation criteria we used software previously developed by some of our colleagues that is able to robustly recognize human faces within arbitrary images [15, 16]. This allowed us to create an image level evaluation criteria of *face affinity* which scores images containing faces highly, and other images poorly. The face detection routines also identify the location and bounds of faces so

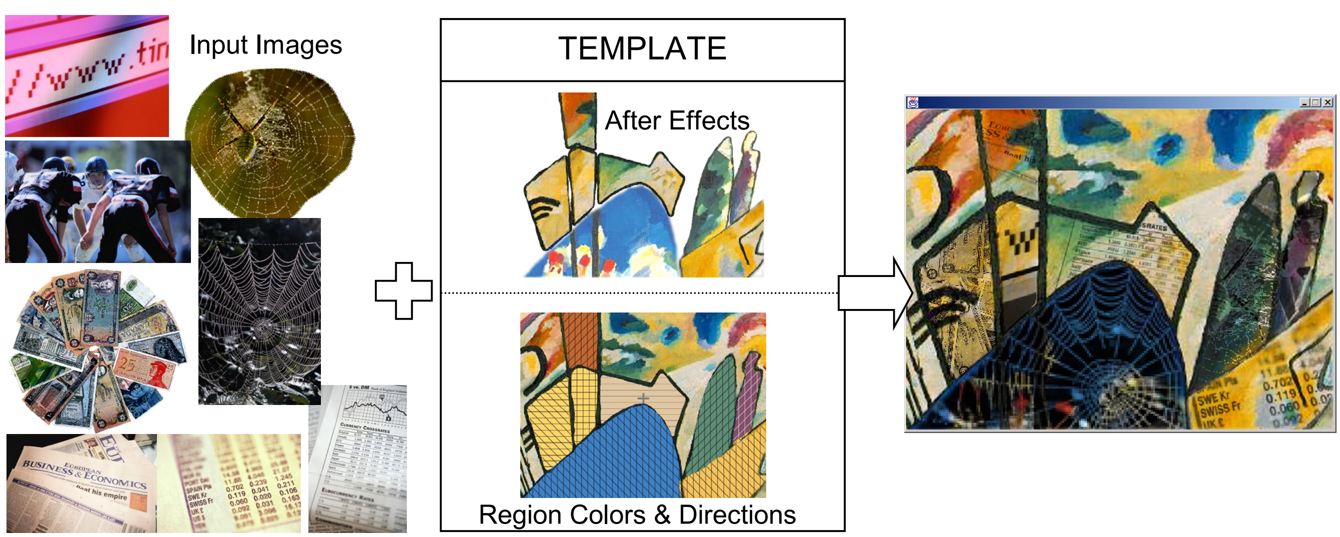

Figure 6. Example Images, Aesthetic Template, and Result

that they may be counted, and so that centers of single faces can be approximated for use in after-effects such as scaling.

THE ARTIST'S INTERFACE

While a general optimization framework provides a number of advantages from the point of view of flexibility and extensibility, the specifications that drive it may also contain a large number of individual criteria that need to be established and parameters that need to be set. Key to the success of the Kandinsky system is its ability to capture and reuse the aesthetic expression of artists. However, this expression normally comes in a fairly fluid form such as sketching or painting. This fluidity will be significantly impeded if the artist is asked to express their work as a list of evaluation criteria, parameters, and weighting factors instead of forms more familiar to them.

To attempt to address this potential mismatch, we have used the region as the central organizational component of our aesthetic templates. From the artist's point of view, regions can be thought of in much the same terms as other constructs they are used to working with in drawing or painting programs – they can be considered primarily in spatial and visual terms. They are organized into the same layer structure that is now common in many painting and illustration programs. Overall, they are defined spatially – covering a particular area of the overall composition (defined internally with a alpha channel mask which might include translucency). Further, parameters needed by many of the primary evaluation criteria can be defined via images. For example, color, direction, and texture matching can all be defined as matches against the properties of the image contained in the region. This allows the majority of the properties of a typical region to be defined simply by drawing an image of the region in a layer in the same way that images are created with the artist's normal tools.

However, some information specifying a region is not visual. This includes the specification of evaluation criteria to be applied or ignored, the weights to be used in combining the scores from criteria, and various semantic criteria (such as an affinity for faces, or limits on the number of images to be associated with a region). At present, we have artists create templates in whatever (layered) tool they normally use, and communicate non-visual criteria to us separately (as annotations on the images), which we enter in tables. This approach, while workable for the artists, clearly is not a viable long-term solution. However, it has shown us that fairly course grained settings of weights (e.g., ratings such as "ignore", "low priority", "medium priority", or "high priority" for each criteria) are generally sufficient.

To allow the artist to directly specify weights and other non-visual attributes in the future, we plan to use a metaphor of *property paints*. Within this metaphor, paints will have colors and patterns much like a conventional painting program. In addition, each paint may contain additional information regarding other properties of a region. For example, paint may contain a factor enabling a particular evaluation criteria such as face affinity. When this *face paint* is applied to a region, the face affinity evaluation criteria becomes enabled for the region. The more area of the region the paint covers, the higher the weighting for that criteria.

We can easily extend this painting metaphor to support facile manipulation of weights and other parameters by supporting dilution and mixing of paints. For example, blue

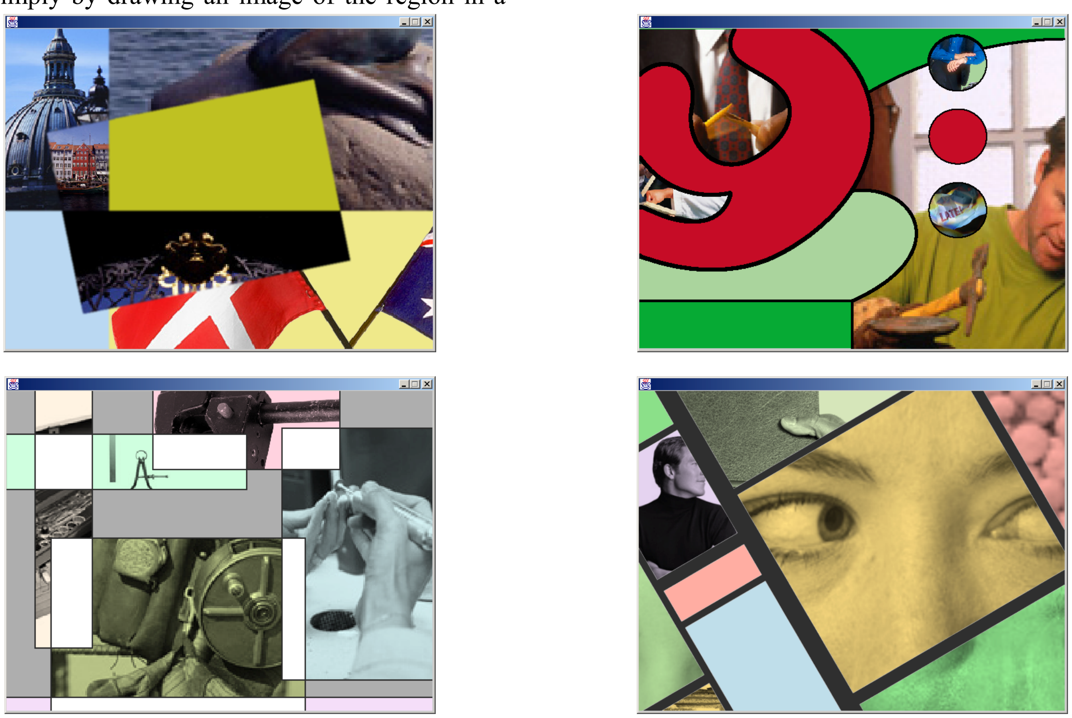

Figure 7. Example Results

paint could be mixed equally with "horizontal" paint to create regions that used color matching and image direction matching criteria equally weighted.

Paint mixing also provides a mechanism by which visual properties of a region can be specified together with non-visual properties. For example, "blue-tint after-effect" paint (which specifies an after-effect of applying a blue hue to the luminance values of a composed region) could be mixed with "face paint", "single image" paint and "centering" paint to create a region that had an affinity for including a single face near its center of mass, then applied the "blue-tint" after-effect.

Overall, the metaphor of using paint to specify properties of regions should provide an intuitive and familiar mechanism for artists to fully define regions. An interesting research issue will be how to best portray to the artists the likely effects of a given paint mixture. At present, the optimization mechanism that causes the criteria implied by a paint to have effect does not run quickly enough for interactive use. However, it may be possible to use a fixed set of example images along with pre-processing and approximation of effects to create approximations of paint actions in real-time. (We can also expect that the optimization algorithm will run in near real time after a few of the CPU speed doublings predicted by Moore's law.)

Figure 7 provides several additional examples of results produced by the Kandinsky system.

CONCLUSION

As computing continues to move from being a highly specialized and expensive tool used by professionals, to a ubiquitous part of most people's every day lives, the need to consider aesthetics, or more generally desirability, in the design of interactive systems will increase. To explore what happens when these criteria are foremost, this paper has turned the normal imperative for information displays upside down. It has considered how technology might be used to create displays that are first aesthetically interesting, and second able to convey information.

ACKNOWLEDGMENTS

We would like to extend special thanks to Henry Schneiderman and Takeo Kanade for making their face recognition software available for this project, and to Doug Cutting for making his Lucene indexing and search package available on the web. We would also like to thank Joonhwan Lee for lending his talents in preparing several of the aesthetic templates used for examples. This work was supported in part by DARPA, the National Science Foundation under Grant IIS-9800597, and the first author's National Science Foundation Graduate Research Fellowship.

REFERENCES

[1] Cohen, H., "The Further Exploits of AARON, Painter", *Stanford Electronic Humanities Review*, v4, n2, 1995 (see: http://www.stanford.edu/group/SHR/4-2/text/cohen.html).

[2] Cutting, D., *The Lucene Search Engine*, Web page: http://www.lucene.com/.

[3] Dondis, D., *A Primer of Visual Literacy*, MIT Press, 1973 (reprinted 2000).

[4] Greenberg, S., and Rounding, M., "The Notification Collage: Posting Information to Public and Personal Displays", *CHI Letters (Proceedings of the SIGCHI Conference on Human Factors in Computing Systems)*, v3, n1, pp. 514 – 521, March 2001.

[5] Heiner, J., Hudson, S., and Tanaka, K., "The Information Percolator: Ambient Information Display in a Decorative Object", *CHI Letters (Proceedings of the ACM Symposium on User Interface Software and Technology)*, v1, n1, pp. 141-148, November 1999.

[6] Hemera Technologies Inc., *Hemera Home Page*, Web page: http://www.hermera.com.

[7] Hofstadter, D. and McGraw, G., "Letter Spirit: An Emergent Model of the Perception and Creation of Alphabetic Style", Technical Report 68, Center for Research on Concepts and Cognition, Indiana University, 1993.

[8] Kandinsky, W., "Composition IV", 1911, (see: http://www.oir.ucf.edu/wm/paint/auth/kandinsky/kandinsky.comp-4.jpg).

[9] Kerne, A., "Collage Machine: Temporality and Indeterminacy in Media Browsing via Interface Ecology" *ACM SIGCHI Conference on Human Factors in Computing Systems Extended Abstracts*, pp. 297-298, March 1997. (see: http://www.mrl.nyu.edu/~andruid/ecology/collageMachine)

[10] Kosslyn, S., *Principles of Graph Design*, W.H. Freeman and Company, 1994.

[11] Mandelbrot, B., "How Long is the Coast of Great Britain, Statistical Self Similarity and Fractional Dimension", *Science*, v155, pp. 636-638, 1967.

[12] Ojala, J., Pietikainen, M, and Harwood, A., "A Comparative Study of Texture Measures with Classification Based on Feature Distributions", *Pattern Recognition*, v 29, pp. 51-50, 1996.

[13] *PhotoDisc: Royalty-free Stock Photography*, Web page: http://www.photodisc.com.

[14] *The Random Collage Generator*, Web page: http://www.albany.net/~egon/.

[15] Schneiderman, H., "A Statistical Approach to 3D Object Detection Applied to Faces and Cars", Doctoral Dissertation, Robotics Institute, Carnegie Mellon University (Available as Robotics Institute Technical Report 00-06), May, 2000.

[16] Schneiderman, H., and Kanade, T., "A Statistical model for 3D Object Detection Applied to Faces and Cars", *IEEE Conference on Computer Vision and Pattern Recognition*, June, 2000.

[17] *Slashdot: News for Nerds, Stuff That Matters*, Web page: http://www.slashdot.org.

[18] Tuceryan, M., and Jain, A., "Texture Analysis" in *The Handbook of Pattern Recognition and Computer Vision (2nd Edition)*, C. Chen, L. Pau, and P. Wang (eds), pp. 207-249, World Scientific Publishing, 1998.

[19] Tufte, E., *Envisioning Information*, Graphics Press, 1990.

[20] Zawinski, J. *WebCollage*, Web page: http://www.jwz.org/webcollage/.

[21] Zhao, Q. and Stasko, J., "What's Happening? The Community Awareness Application", *ACM SIGCHI Conference on Human Factors in Computing Systems Extended Abstracts*, pp. 253-254, April 2000.