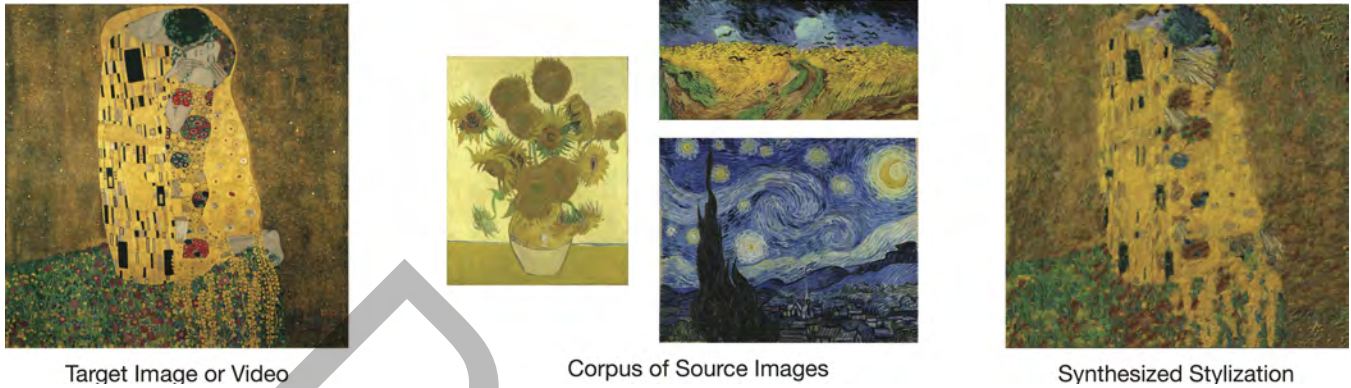# Corpus-Based Visual Synthesis: An Approach for Artistic Stylization

Parag K. Mital[1]*, Mick Grierson[1], Tim J. Smith[2]
[1]Goldsmiths, University of London
[2]Birkbeck, University of London

Figure 1: *Klimt's "The Kiss" is synthesized using 3 images of Van Gogh paintings to produce the result on the right. Best viewed in color at 400%. Images representing faithful reproductions of Gustav Klimt and Van Gogh sourced from Wikimedia Commons are public domain.*

Abstract

We investigate an approach to the artistic stylization of photographic images and videos that uses an understanding of the role of abstract representations in art and perception. We first learn a database of representations from a corpus of images or image sequences. Using this database, our approach synthesizes a target image or video by matching geometric representations in the target to the closest matches in the database based on their shape and color similarity. We show how changing a few parameters of the synthesis process can result in stylizations that represent aesthetics associated with Impressionist, Cubist, and Abstract Expressionist paintings. As the stylization process is fast enough to work in real-time, our approach can also be used to learn and synthesize the same camera image, even aggregating the database with each new video frame in real-time, a process we call "Memory Mosaicing". Finally, we report the user feedback of 21 participants using an augmented reality version of "Memory Mosaicing" in an installation called Augmented Reality Hallucinations, where the target scene and database came from a camera mounted on augmented reality goggles. This information was collected during an exhibition of 15,000 participants at the Digital Design Weekend at the Victoria and Albert Museum (co-located during the London Design Festival).

CR Categories: I.3.3 [Computer Graphics]: Picture/Image generation—Display Algorithms I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction Techniques J.5 [Computer Applications]: Arts and Humanities—Fine Arts

Keywords: memory mosaicing, augmented reality hallucination

1 Introduction

Despite its apparent precision, our perception of reality is not representative of the way that we see. For instance, the light coming to our eyes is distorted, upside-down, and constantly disrupted with each movement of the eye. How can this noisy process ever constitute our experience of the visual world? Numerous theories have argued that in order to perceive the world as a continuous and richly detailed one, our vision system must use abstracted representations of the world [Marr 1982]. It is argued that these representations are created by grouping together coherent visual features that resemble abstract forms - such as geometrical primitives. Grouping such primitives together eventually leads to the formation of semantic representations such as objects. Importantly, the representations used in vision are not necessarily what we perceive, but are what we use in order to help us perceive. As a result, these representations are likely to remove details that are unimportant to a person's ongoing task while making other details more explicit.

Artists are well aware of the role of representation in perception. By leaving out particular details from a visual scene and accentuating others, they are able to direct a viewer's attention within a visual medium, influencing their perception [Haeberli 1990; Zimmer 2003]. Picasso once famously said, "I paint forms as I think them, not as I see them" [Hughes 1991]. As one of the pioneers of Cubism, Picasso wanted to represent the fact that our perception of an object is based on all possible views of it. He did so by compressing all views of an object into a synthesized one built using abstracted shape primitives. Other movements in art can also be

*e-mail:parag@pkmital.com

characterized as utilizing representations formed through geometrical primitives. In Impressionist painting, these forms are often described by a dense number of short and visible brush strokes. In Abstract Expressionist painting, the primitives are again dense, though tended to be of much larger strokes in an attempt to abstract away as much detail of a real scene as possible.

In this paper, we investigate an approach to the artistic stylization of photographic images and videos through the use abstracted shape representations. The representations that are built by this method can be varied in size and density using a process that allows the user to manipulate parameters in real-time. Our system first learns a database of representations from a corpus of images. It then synthesizes a target image or video by matching geometric representations in the target to the closest matches in the database. We show how changing the parameters of the synthesis process results in stylizations that represent aesthetics associated with Impressionist, Cubist, and Abstract Expressionist paintings. As the stylization process is fast enough to work in real-time, this approach can also be used to learn and synthesize the same camera image, even aggregating the database with each new video frame in real-time, a process we call "Memory Mosaicing". Finally, we report the feedback of 21 participants using an augmented reality version of "Memory Mosaicing" in an installation called Augmented Reality Hallucinations, where the target scene and database came from a camera mounted on augmented reality goggles.

2 Related Work

Artistic stylization has seen significant advances over the last 14 years. Kyprianidis recently surveyed the field in [Kyprianidis et al. 2012]. The field began as filtering and clustering algorithms were applied to images, accentuating regions within an existing image to produce aesthetics associated with different styles (e.g., for Pointillism [Yang and Yang 2006; Seo and Yoon 2010]; for cartoonization [Wang et al. 2004b]; for oil and watercolor [Meier 1996; Hertzmann and Perlin 2000; Bousseau et al. 2007; Gooch et al. 2002]; for Impressionism [Litwinowicz 1997; Hertzmann 1998]). More recent approaches focused on using user-guided segmentation, where the user manually labels key frames with strokes defining how the frame is stylized (e.g. [O'Donovan and Hertzmann 2012]) or uses eye-movements in deciding which aspects of a photo are most salient [DeCarlo and Santella 2002].

Hertzmann's seminal work in Image Analogies [Hertzmann et al. 2001] presented a branch from the aforementioned approaches by allowing control of the stylization process through choosing a pair of example images. By finding the patterns associated with an existing stylization of an image A to another image A', a user could then stylize a target image B by analogy into B' (later extended to include analogies between curved strokes [Hertzmann et al. 2002]). In the same year, [Efros and Freeman 2001; Liang et al. 2001] also developed methods in texture transfer and patch-based sampling, where existing image material was used to synthesize textures of arbitrary sizes. These methods were later extended in [Wang et al. 2004a], where a user specified small blocks in an example painting that represented the style to recreate. These blocks were then synthesized along computed paint strokes in the target image using an efficient hierarchical texture synthesis method. Though Wang's approach and even more recent methods (e.g., [Guo et al. 2006]) produces impressive results, it also relies on user interaction to select the representative patches expressing an artistic style. Further, the aforementioned work in texture transfer as well as more recent approaches (e.g., [Lee et al. 2010]) all rely on a single source image in order to transfer the style of the texture, meaning the range of stylizations possible are constrained to the information contained in a single image. In this paper, we develop an approach that does not require the user to manually label any regions and that is not confined to a single example image while still affording a range of possible styles.

Our approach, corpus-based visual synthesis (CBVS), synthesizes a target image/video using existing pre-defined visual content. As a result, it is also borrows methods from dictionary-based approaches ([Zeng et al. 2009; Healey et al. 2004]), though our approach does not focus on developing strokes from expert training as we automatically segment a corpus of user chosen images. It also shares methodology with collage/mosaic-based work (e.g. [Kim and Pellacini 2002; Orchard and Kaplan 2008; Huang et al. 2011; Miller and Mould 2012]), allowing a user to work with a period of an artist's work or entire videos, for example. Though these approaches are targeted for collage/mosaic-based purposes rather than artistic stylization, [Huang et al. 2011] describes an approach that is also motivated by an artist making use of collage. Their approach produces what they call "Arcimboldo-like" collages in the style of 18th century painter Giuseppe Arcimboldo, relying on user strokes to segment the images used. In contrast, CBVS is aimed towards producing a range of possible artistic stylizations through changing a few simple parameters. Further, as segmentation happens without requiring user-selected patches or strokes, CBVS is also suitable for producing stylization of videos, unlike the very impressive though slow approach (15 minutes for a 300 x 400 pixel image) reported in [Chang et al. 2010].

3 Corpus-based Visual Synthesis Framework

CBVS begins by first aggregating all frames from a user chosen corpora of images, $\mathbf{C} = \{C_1, C_2, ..., C_N\}$, containing N total candidate images. We aim to use the content solely from this corpus to artistically stylize a target image or video, $\mathbf{T} = \{T_1, T_2, ..., T_M\}$, containing M total frames. We develop a rendering procedure for image and video-based targets where parameters of the synthesis can be changed interactively. To begin, we describe detection, tracking, description, matching, and synthesis of the abstracted shape representations. We then describe parameters influencing each of these steps before showing our results in Section 5.

3.1 Detection

For both the candidate and target frames, we aim to detect abstracted shape primitives described by coherent image regions. For this purpose, we make use of maximally stable color regions (MSCR) [Forssén 2007]. The algorithm described in [Forssén 2007] successively clusters neighboring pixels with similar colors described by multiple thresholds of a distance measure which takes into account the inherent camera noise and the probability distribution of each RGB color channel. Regions are denoted as maximally stable if they do not grow larger than a minimum margin for certain number of time-steps. Previous techniques employing posterization, filtering, or watershed have had to apply their algorithm at multiple scales in order to discover regions that are superimposed or overlapped, increasing their computational complexity. MSCR has the benefit over these previous techniques as it provides an implicit ordering of superimposed regions discovered through successive time-steps of the clustering algorithm. Further, it allows us to prune regions by restricting their area to a range of minimum and maximum sizes. In Section 4.1, we discuss these parameters in greater detail in relation to the styles they can produce. We use MSCR to detect the set of all regions in each candidate and target frame, denoted as $\mathbf{R_C} = \{R_1, R_2, ..., R_{N_C}\}$ and $\mathbf{R_T} = \{R_1, R_2, ..., R_{N_T}\}$ where N_C is the number of regions

detected in all candidate frames and N_T is the number of target regions.

3.2 Tracking

It is often desirable to produce temporally coherent stylizations, meaning if a region within a target video frame has not moved, it is not re-stylized. This is especially the case in noisy or compressed videos, where artifacts may appear that should not be stylized. One approach would be to track regions using a GPU-based Optical Flow measure. This would likely produce reasonable temporal coherence without sparing real-time interaction. However, we simply follow [Hertzmann and Perlin 2000] in using the flicker for detecting the change in the original target video, as this approach is fast and easy to compute. Let the flicker for a pixel at location (i, j) be described by:

$$f(i, j) = I_t(i, j) - I_{t-1}(i, j) \qquad (1)$$

where I is the image luminance at time t. Then, if the flicker at the region's centroid, $f(C_{R_i})$, between the current and previous frame is greater than a threshold, $threshold$, we remove the region from the set of detected regions to synthesize:

$$R_T = \{R_i \mid f(C_{R_i}) > threshold, \forall i = 1...N_T\} \qquad (2)$$

3.3 Description

We form a descriptor comprised of shape and color values. The shape descriptor for each region, d_{R_i}, is composed of the normalized central moments up to order 2. The average color of the region is converted from RGB to the 3-channel CIELAB color space, L, a^*, b^*. These form the final descriptor:

$$d_{R_i} = \left(\mu_{00}, \eta_{11}, \eta_{20}, \eta_{02}, L, a^*, b^*\right) \qquad (3)$$

where μ_{ij} is the central image moment of order i and j, i.e. μ_{00} is simply the area, and η_{ij} is the normalized central image moment computed as:

$$\eta_{ij} = \frac{\mu_{ij}}{\mu_{00}^{\left(1 + \frac{i+j}{2}\right)}} \qquad (4)$$

Centralizing the moments allows us to compare regions with translation-invariance, while normalizing the first and second order moment allows us to compare regions with scale-invariance. We include the area as the first term as this ensures regions are not distorted too much when matching. Further, employing CIELAB allows us to define the region in a color space where we can then use perceptual metrics for matching. We describe this metric in greater detail in the next section.

3.4 Matching

We match each region in the target to its nearest neighbors in the database using a metric combining distances from each region's shape and color, $d_s(R_t, R_c)$ and $d_c(R_t, R_c)$, respectively:

$$d(R_t, R_c) = d_s(R_t, R_c) + d_c(R_t, R_c) \qquad (5)$$

The shape distance is simply computed as the absolute difference between the first and second order normalized central image moments of each region (i.e. the first four components of the descriptor). For the color distance, we make use of the official CIE color-difference equation, CIEDE2000, which provides reliable color discrimination with interactive terms for lightness, chromaticity, and hue weighting [Luo et al. 2001]. This difference formula has been shown to be more perceptually accurate at determining the difference between colors than previous methods employing linear difference using RGB or LUV color values, as it is based on empirical evidence of perceived color difference. For our tests, we use the default parameters described in [Luo et al. 2001] for the weighting terms.

3.5 Synthesis

To ensure regions are drawn from their background to the foreground, we synthesize each target region in order from the largest to smallest area sizes. In contrast to methods that place brush strokes based on the stroke direction at each pixel on the medial axis (e.g.,[Wang et al. 2004a]), we find the affine geometric transform describing the transformation from R_{C_i} to R_{T_i}. This can be described by a translation, rotation, and scaling. The translation component is simply the difference in each region's centroid. The rotation can be found using the central image moments:

$$\Theta = \frac{1}{2} * \arctan \frac{2 * \frac{\mu_{11}}{\mu_{00}}}{\frac{\mu_{20}}{\mu_{00}} - \frac{\mu_{02}}{\mu_{00}}} \qquad (6)$$

Finally, scaling is simply the ratio of the target to candidate region's bounding box. This process has the benefit of being very fast using graphics hardware as it can be computed by a single matrix multiplication. Each region is then layered above the previous one before creating a synthesized image. In image-based stylization, multiple syntheses created with changing parameters can be blended together to create more detailed and expressive styles which may require many "layers" of "paint". We discuss these parameters in greater detail in the next section.

4 Parameters

Parameters influencing the region detection algorithm are set independently for the corpus and the target, as their function differs.

4.1 Corpus Parameters

For the corpus, we define the *timesteps*, *minimum region area*, and *maximum region area* of the detected regions. We use a set of parameters that learns the widest range of possible regions covering both small and large regions. In some cases, as in more abstract styles, it may be desirable to learn a very small number of regions, limiting the range of expressiveness to a few possible primitives. As the timesteps parameter influences the number of evolutions allowed in the MSCR algorithm, the higher this number, the more regions will be discovered. Similarly, lowering the minimum region size and increasing the maximum region size reduces the number of region that are pruned. In our tests, we found a single set of parameters to be sufficient for defining a varied corpus: 100 for the timesteps, 35 pixels for the minimum region area, and 50% of the image's size for maximum region area.

When learning a corpus from many images, we restrict learning regions that are within a distance threshold (using Equation 3.4) of all regions in the existing database. For our examples, we set this parameter to 50. This value is low enough to include many regions, though high enough to avoid detecting duplicate regions. A higher number for this parameter will lead to very discriminative regions. In our tests, when setting this number higher, we found that our corpus had less variety of regions to synthesize from, leading to stronger shape or color mismatches.
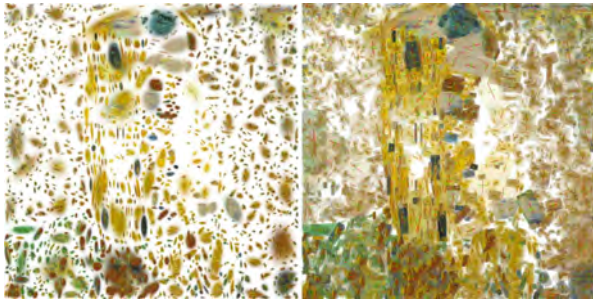
Figure 2: *Using the target image and database shown in Figure-1, we show an example stylization with (first image) and without (second image) spatial blending. We also draw the region's orientation depicted by red/green axes in order to better show the regions (best viewed in the color manuscript at 200%).*
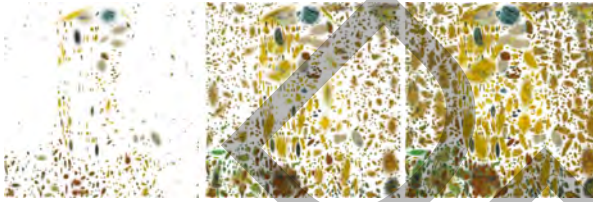
Figure 3: *Using the target image and database shown in Figure-1, the timesteps are increased over time. This allows the user to detect more regions and develop a denser and higher contrast stylization.*

4.2 Target Parameters

For the target, we allow the user to interactively define a few parameters affecting the output stylization.

- *Spatial blending*: Allows the user to use feathered elliptical regions instead of rectangular ones (see Figure-2). When stylizing finer details of an image, this parameter is very useful for removing hard edges produced by rectangular regions.

- *Timesteps*: Increasing this produces more regions, making the image denser (see Figure-3). As well, this will also produce more regions that coincide with each other. As a result, when synthesizing with a high number for the timesteps, the result resembles an overpainting effect. For styles that require many "layers" of "paint", we use a higher number for the timesteps. When used in combination with blending, increasing this can also increase the contrast.

- *Minimum region size*: This parameter determines the minimum allowed region size for synthesis. Setting this number very low (e.g. below 100 pixels) produces styles more similar to Impressionism, as many small regions are detected (see Figure-4).

- *Maximum region size*: Similar to the minimum region size parameter, this parameter determines the largest allowed region size. Generally setting this number as high as possible will be sufficient. However, it may be desirable to interactively change this parameter over time, allowing for large regions to be drawn at first, then only allowing smaller ones.

- *Temporal blending*: Uses alpha blending to composite regions over time (see Figure-6). Together with an increased number of timesteps, this parameter can be used to change the contrast of the overall image (as shown in Figure-7).

- *Motion tracking*: Allows regions to be drawn only if their de-

Figure 4: *Using the target image and database shown in Figure-1, the minimum region size is decreased over time, allowing the user to detect smaller regions and produce finer detailed stylizations.*

Figure 5: *Using the target image and database shown in Figure-1, the blending radius is increased over time. This parameter influences the overall size of the drawn regions. Setting this number smaller can help to produce finer details on top of existing layers, often associated with both Impressionist and Abstract Expressionist styles.*

tected motion is higher than a fixed threshold. For our experiments, we set this number to 5.

- *Blending radius*: Influences the feathering radius of the detected region (see Figure-5). Normally, each detected region is matched to one in the database and then through an affine transformation placed where the detected region was using the same scale and rotation. However, it may be desirable to change the scale of this region using the blending radius to produce different effects. When scaling this region down, a user confines drawing to only small regions being painted, often produces styles associated with Abstract Expressionism.

For image-based targets, the aforementioned parameters effect the frame-to-frame compositing, meaning the same image is rendered over itself. For video-based targets, however, only a single iteration is used for each frame, as much of the information required for building styles requiring more detailed composites can be extracted over the first 1 or 2 frames. We demonstrate how these parameters can influence a wide range of stylizations in the next section.

5 Results

We use the presented framework to produce artistic stylizations of photo-realistic images and videos. In this section, we show our results in image-based stylization using a landscape, abstract, and painterly scene. We then show how the same framework can be used with video targets, including an abstract and portrait video. As well, we show a particular case where the source material is aggregated from a live-stream of the target, i.e. the source and target are the same, a process we call "Memory Mosaicing". Finally, we present an augmented reality version of "Memory Mosaicing" including feedback collected from 21 participants of an installation at the Victoria and Albert Museum in London.

5.1 Image: Landscape

In Figure-8, we synthesize a landscape photo of cows grazing using Expressionist painter Paul Klee. We turn off spatial blending and

Figure 6: *Using the target image and database shown in Figure-1, we increase the temporal blending factor. This influences the opacity of every region drawn.*

Figure 7: *Using the target image and database shown in Figure-1, we use temporal blending as well as decreasing minimum region size and increased timesteps to begin to produce the final synthesis.*

use a small value for the minimum region size. We also allow the maximum region size to be very large. This results in a relatively smaller region being matched to the sky and stretched to fill the top-half of the image. The synthesized region happens to look like a rainbow, though the original region itself was very abstract (see the corner piece of Paul Klee's "Rote und weisse Kuppeln").

5.2 Image: Abstract

In Figure-9, we synthesize a close-up picture of a blanket using Klimt's The Kiss. The target this time is very abstract and we will not need to synthesize parameters that force an abstract quality rendering such as large region sizes. As such, we allow the minimum region size to be very small producing more details, though retaining a style associated with Abstract Expressionism.

5.3 Image: Painterly

We demonstrate how CBVS can stylize existing painterly images into other styles. In the teaser graphic in Figure-1, we use three paintings by Van Gogh to stylize Klimt's The Kiss. Here, we set the minimum region size to be small, allowing finer details and smaller brush strokes, and allow the timesteps to be high as we want to bring out as much contrast as possible.

In Figure-10, we try synthesizing Van Gogh's The Bedroom using 3 images of Monet's Water Lilies series. Here, we ensure we detect many small regions by increasing the timesteps and setting the minimum region size to be very small. Further, we turn on spatial blending as we decrease the minimum region size, as we want to avoid rendering any strong edges, retaining an Impressionist quality.

5.4 Video: Portrait

Two examples in video-based stylization are presented: one of a subject rowing a boat and another of abstract imagery. In Figure-11, we can see 4 frames taken from a video stylization. We use the same corpus as in Figure-8 and allow the minimum region size to be very small, resulting in a more Expressionist style. The first frame is not as composed as the later frames, as there will have only been 1 frame of compositing. As a result, the first frame in video-based

Expressionist stylization may not be a consistent style with its later frames.

5.5 Video: Abstract

In Figure-12, we stylize a video using the same corpus as in Figure-8 and set the minimum region size to be very large. Thus, instead of producing an Expressionist style as in Figure-11, less details are synthesized resulting in a more abstract style. The first frame in this video does not necessarily require more than 1 iteration as it is synthesizing very large regions that often also overlap.

5.6 Memory Mosaicing

The artistic stylization process can be used in a real-time context without an explicit corpus. In this case, we aggregate representations learned from the ongoing stream of target frames. Parameters are generally set by the user interacting with the process, or contained to a single preset. In particular, restricting the total number of representations as first-in-first-out queue allows the process to continue in real-time with a simple linear search index. In the examples shown in Figure 13, we show two example outputs from the same camera stream. In the left image, we aim for large region sizes and low timesteps, resulting in a more abstract style, reminiscent of Cubist style paintings. In the right example, we allow higher timesteps and only small region sizes, resulting in a more expressive style similar to paintings in Abstract Expressionism.

5.7 Augmented Reality Hallucination

An interesting case of "Memory Mosaicing" is when a participant can actively explore a visual scene. By using augmented reality goggles, we allowed participants to explore their environment through the our proposed stylization process during an exhibition called "Augmented Reality Hallucinations" held at the Victoria and Albert Museum in London. Participants were invited to wear the goggles where two small CRT screens presented the same output of a "Memory Mosaicing" of a single camera mounted on the goggles right eye that faced the scene in front of them (see Figure 14). As the only user interaction was in exploring a scene, a single preset was defined based on large region sizes and low number of timesteps, as shown in the left column in Figure 13.

Participants were also invited to give quantitative and qualitative feedback on their experience. The summary of the quantitative feedback is shown in Figure 15. On the feedback form, when participants were asked, "Did this experience make you think of anything you had seen or heard before?", three participants made references to their experiences on hallucinogens and two to dreams. Also of note in the qualitative feedback was references to art styles such as, "It reminded me of Francis Bacon's Figurative style" and "The movement was Impressionistic, almost painterly". When asked, "What did you dislike most about the experience?", of note were the responses, "Would have liked more depth in colour", "Not sure what I was seeing at first with the goggles", and "Hard to understand how it works." The lack of understanding of the process may also be revealed in the quantitative analysis in the second bar of the graph. However, on average, this number is still quite high across participants, though there is also no baseline to compare to.

6 Discussion and Future Works

We have presented a framework for producing artistic stylizations of images or videos. A corpus of image material is automatically segmented, defining the possible strokes effecting the possible colors and textures in the stylization. Using a simple set of parameters, we have shown that many stylizations of a target image or video

Figure 8: *A landscape picture of cows grazing is synthesized using 13 images of Expressionism painter Paul Klee to produce the image on the right. Corpus images representing faithful reproductions of Paul Klee were sourced from Mark Harden's Artchive. Photo of cows taken by the author.*
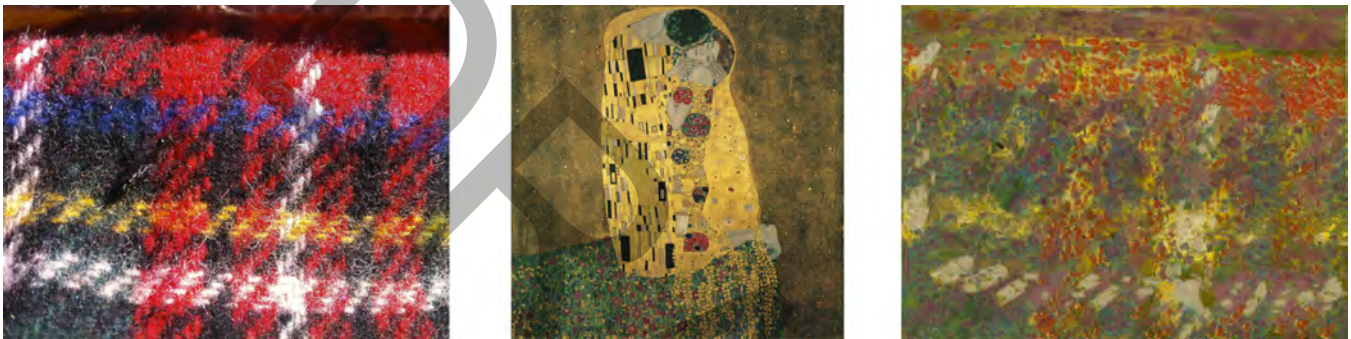
Figure 9: *A close-up picture of a blanket is synthesized using Klimt's The Kiss to produce the image on the right. Best viewed in the color manuscript at 200%. Images representing faithful reproductions of Gustav Klimt sourced from Wikimedia Commons are public domain. Photorealistic scene of blanket taken by the author.*

are possible, ranging from Impressionism, Expressionism, and Abstract Expressionism. By allowing the interactive refinement of an image's stylization, we allow the user to experiment with a range of stylizations through simple parameters. This interactive refinement affords compositing, the ability to blend together stylizations from different parameters over time. We also demonstrate the extension of this framework to video-based stylization using simple motion tracking. As in image-based stylization, the user can influence the stylization through the same set of parameters in real-time to interactively refine the stylization.

The extension of video-based stylization is also particularly suited for real-time contexts as shown in "Memory Mosaicing", where a database is aggregated from learning representations in a target frame over time. Extending this case to an augmented reality setting, a participant of this system can actively view their world, creating a hallucinogenic experience, as validated by a number of participants during an exhibition at the Victoria and Albert Museum. However, the feedback from this installation also revealed a lack of understanding in how the process works.

A number of issues could be addressed in future versions. For instance, synthesized regions with poor shape matches can be heavily distorted in a resulting synthesis. In these cases, it is likely that the database did not include any other matches with more similar shapes, or the shape descriptor had been weighted too low. As well, the speed of the synthesis in a real-time context can be greatly im-

proved with other search methods such as tree or hash-table based indexes. As well, our approach to addressing the temporal coherence of the resulting stylization may be improved with investigating incorporating more recent models of optical flow, keyframe detection, and possibly spatiotemporal detection of representations rather than purely spatial ones.

References

BOUSSEAU, A., NEYRET, F., THOLLOT, J., AND SALESIN, D. 2007. Video watercolorization using bidirectional texture advection. *ACM SIGGRAPH 2007 papers on - SIGGRAPH '07*, 104.

CHANG, Y., SAITO, S., UCHIKAWA, K., AND NAKAJIMA, M. 2005. Example-Based Color Stylization of Images. *ACM Transactions on Applied Perception 2*, 3 (July), 322–345.

CHANG, I., PENG, Y., CHEN, Y., AND WANG, S. 2010. Artistic Painting Style Transformation Using a Patch-based Sampling Method. *Journal of Information Science and Engineering 26*, 1443–1458.

DECARLO, D., AND SANTELLA, A. 2002. Stylization and abstraction of photographs. *ACM Transactions on Graphics 21*, 3 (July), 1–8.

EFROS, A., AND FREEMAN, W. 2001. Image quilting for texture synthesis and transfer. *SIGGRAPH 2001: Proceedings of the 28th annual conference on Computer graphics and interactive techniques.*.

Figure 10: *Van Gogh's "The Bedroom" is synthesized using 3 images of Monet paintings to produce the image on the right. Images representing faithful reproductions of Van Gogh and Claude Monet sourced from Wikimedia Commons are public domain.*

Figure 11: *Left: 4 frames from a target video; Right: Stylization using the same corpus as in Figure-8. We aim to synthesize with greater expression and less abstraction, and allow the minimum region size to be very small. Best viewed in the color manuscript at 200% or in the video online. Photos by the author.*

Figure 12: *Left: 4 frames from a target video; Right: Stylization using the same corpus as in Figure-8. Here we aim to stylize with greater abstraction than in Figure-11, and set the minimum region size to be fairly large. Best viewed in the color manuscript at 200% or in the video online. Photos by the author.*

FORSSÉN, P.-E. 2007. Maximally stable colour regions for recognition and matching. *Computer Vision and Pattern Recognition 2007, (CVPR07)..*

GOOCH, B., COOMBE, G., AND SHIRLEY, P. 2002. Artistic vision: painterly rendering using computer vision techniques. In *NPAR '02 Proceedings of the 2nd international symposium on Non-photorealistic animation and rendering*, 83.

GUO, Y.-W., YU, J.-H., XU, X.-D., WANG, J., AND PENG, Q.-S. 2006. Example based painting generation. *Journal of Zhejiang University SCIENCE A 7*, 7 (June), 1152–1159.

HAEBERLI, P. 1990. Paint by numbers: abstract image representations. *ACM SIGGRAPH Computer Graphics 24*, 4 (Sept.), 207–214.

HEALEY, C. G., TATEOSIAN, L., ENNS, J. T., AND REMPLE, M. 2004. Perceptually based brush strokes for nonphotorealistic visualization. *ACM Transactions on Graphics 23*, 1 (Jan.), 64–96.

HERTZMANN, A., AND PERLIN, K. 2000. Painterly rendering for video and interaction. *Proceedings of the first international symposium on Non-photorealistic animation and rendering - NPAR '00*, 7–12.

HERTZMANN, A., JACOBS, C. E., OLIVER, N., CURLESS, B., AND SALESIN, D. H. 2001. Image analogies. *Proceedings of the 28th annual conference on Computer graphics and interactive techniques - SIGGRAPH '01*, 327–340.

HERTZMANN, A., OLIVER, N., CURLESS, B., AND SEITZ, S. M. 2002. Curve analogies. In *EGRW '02 Proceedings of the 13th Eurographics workshop on Rendering*, 233–246.

HERTZMANN, A. 1998. Painterly rendering with curved brush strokes of multiple sizes. *Proceedings of the 25th annual conference on Computer graphics and interactive techniques - SIGGRAPH '98*, 453–460.

HUANG, H., ZHANG, L., AND ZHANG, H.-C. 2011. Arcimboldo-like collage using internet images. *Proceedings of the 2011 SIGGRAPH Asia Conference on - SA '11 30*, 6, 1.

HUGHES, R. 1991. *Shock of the New*.

KIM, J., AND PELLACINI, F. 2002. Jigsaw image mosaics. *ACM Transactions on Graphics 21*, 3 (July).

KYPRIANIDIS, J., COLLOMOSSE, J., WANG, T., AND ISENBERG, T. 2012. State of the'Art': A Taxonomy of Artistic Stylization Techniques for Images and Video. *IEEE transactions on Visualization and Computer Graphics*.

LEE, H., SEO, S., RYOO, S., AND YOON, K. 2010. Directional texture transfer. *NPAR '10 Proceedings of the 8th International Symposium on Non-Photorealistic Animation and Rendering 1*, 212, 43–50.

LIANG, L., LIU, C., XU, Y.-Q., GUO, B., AND SHUM, H.-Y. 2001. Real-time texture synthesis by patch-based sampling. *ACM Transactions on Graphics 20*, 3 (July), 127–150.

LITWINOWICZ, P. 1997. Processing images and video for an impressionist effect. *SIGGRAPH '97 Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, 407–414.

LUO, M. R., CUI, G., AND RIGG, B. 2001. The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color Research & Application 26*, 5 (Oct.), 340–350.

MARR, D. 1982. *Vision: A Computational investigation into the Human Representation and Processing of Visual Information*.

MEIER, B. J. 1996. Painterly rendering for animation. *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques - SIGGRAPH '96*, 477–484.

MILLER, J., AND MOULD, D. 2012. Accurate and Discernible Photocollages. *Computational Aesthetics in Graphics, Visualization, and Imaging*, 115–124.

O'DONOVAN, P., AND HERTZMANN, A. 2012. AniPaint: interactive painterly animation from video. *IEEE transactions on visualization and computer graphics 18*, 3 (Mar.), 475–87.

ORCHARD, J., AND KAPLAN, C. 2008. Cut-out image mosaics. *ACM Transactions on Graphics 1*, 212.

SEO, S., AND YOON, K. 2010. Color juxtaposition for pointillism based on an artistic color model and a statistical analysis. *The Visual Computer: International Journal of Computer Graphics 26*, 6-8 (Apr.), 421–431.

WANG, B., WANG, W., YANG, H., AND SUN, J. 2004. Efficient example-based painting and synthesis of 2D directional texture. *IEEE transactions on visualization and computer graphics 10*, 3, 266–77.

WANG, J., XU, Y., SHUM, H.-Y., AND COHEN, M. F. 2004. Video tooning. In *SIGGRAPH '04 ACM SIGGRAPH 2004 Papers*, ACM Press, New York, New York, USA, 574–583.

YANG, H., AND YANG, C. 2006. A Non-Photorealistic Rendering of Seurat's Pointillism. *Advances in Visual Computing*, 760–769.

ZENG, K., ZHAO, M., XIONG, C., AND ZHU, S. 2009. From image parsing to painterly rendering. *ACM Transactions on Graphics (TOG) 29*, 1.

ZIMMER, R. 2003. Abstraction in art with implications for perception. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences 358*, 1435 (July), 1285–91.
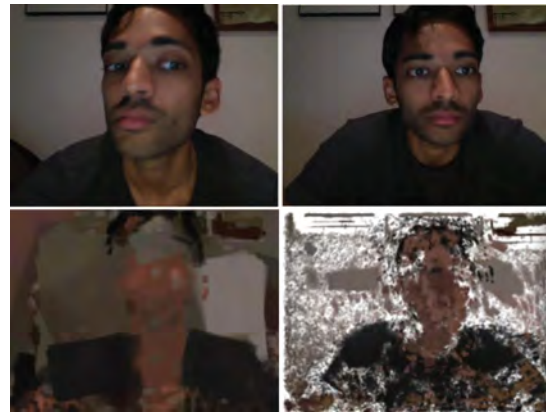
Figure 13: *2 examples of "Memory Mosaicing" showing the input (top) and resulting real-time stylization (bottom). Photos by the author.*

Figure 14: *An exhibition at the Victoria and Albert Museum in London had participants wear Augmented Reality goggles with software running a real-time version of "Memory Mosaicing". Photos by the author.*
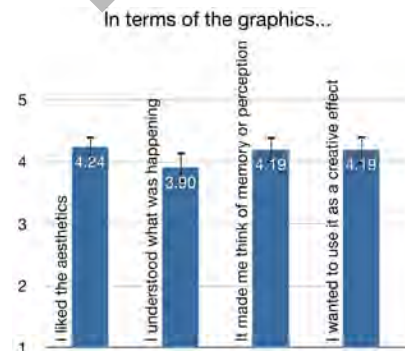
Figure 15: *Results of the "Augmented Reality Hallucination" installation feedback where 21 participants were asked to rate different aspects of the visual synthesis. Error bars depict +/- 1 S.E.*