Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 8, No. 1, August : 2023

TEXT-BASED BINARY IMAGE STYLIZATION AND SYNTHESIS

Senthil Pandian P, Department of Computer Science and Engineering, Solamalai College of Engineering, Madurai, Tamilnadu, India
R Selvabharathi, Department of Mechanical Engineering, AAA College of Engineering and Technology, Virudhunagar, Tamilnadu, India
Vivek V, Department of Computer Science and Engineering, AAA College of Engineering and Technology, Virudhunagar, Tamilnadu, India

Abstract
In this paper, I present a new framework for the stylization of text-based binary images. First, this method stylizes the stroke-based geometric shape like text, symbols and icons in the target binary image based on an input style image. Second, the composition of the stylized geometric shape and a background image is explored. To accomplish the task, I propose legibility- preserving structure and texture transfer algorithms, which pro- gressively narrow the visual differences between the binary image and the style image. The stylization is then followed by a context- aware layout design algorithm, where cues for both seamlessness and aesthetics are employed to determine the optimal layout of the shape in the background. Given the layout, the binary image is seamlessly embedded into the background by texture synthesis under a context-aware boundary constraint. According to the contents of binary images, this method can be applied to many fields. I show that the proposed method is capable of addressing the unsupervised text stylization problem and is superior to state- of-the-art style transfer methods in automatic artistic typography creation. Besides, extensive experiments on various tasks, such as visual-textual presentation synthesis, icon/symbol rendering and structure-guided image inpainting, demonstrate the effectiveness of the proposed method.

Keywords: Texture synthesis, structure synthesis, context- aware, style transfer, image inpainting.

I. INTRODUCTION

STYLE transfer is the task of migrating a style from an image to another to synthesize a new artistic image. It is of special interest in visual design, and has applications such as painting synthesis and photography post-processing. However, creating an image in a particular style manually requires great skills that are beyond the capabilities of average users. Therefore, automatic style transfer has become a trending topic both in academic literature and industrial applications. Text and other stroke-based design elements such as symbols, icons and labels highly summarize the abstract imagery of human visual perceptions and are ubiquitous in our daily life. The stylization of text-based binary images as in Fig. 1(a) is of great research value but also poses a challenge of narrowing the great visual discrepancy between the binary flat shapes and the colorful style image.

Style transfer has been investigated for years, where many successful methods are proposed, such as the non-parametric method Image Quilting [1] and the parametric method Neural Style [2]. Non-parametric methods take samples from the style image and place the samples based on pixel intensity [1], [3],
[4] or deep features [5] of the target image to synthesize a new image. Parametric methods represent the style as statistical features, and adjust the target image to satisfy these features. Recent deep learning based parametric methods [2], [6], [7] exploit high-level deep features, and thereby have the superior capability of semantic style transfer. However, none of the aforementioned methods are specific to the stylization of text- based binary images. In fact, for non-parametric methods, it is hard to use pixel intensity or deep features to establish a direct mapping between a binary image and a style image, due to their great modality discrepancy. On the other hand, text-based binary images

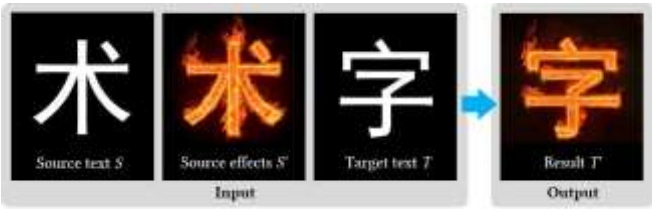
lack high-level semantic information, which limits the performance of the parametric methods.

As the most related method to our problem, a text effects transfer algorithm [8] is recently proposed to stylize the binary text image. In that work, the authors analyzed the high cor- relation between texture patterns and their spatial distribution in text effects images, and modeled it as a distribution prior, which has been proven to be highly effective at text stylization. But this method strictly requires the source style to be a well- structured typography image. Moreover, it follows the idea of Image Analogies [9] to stylize the image in a supervised manner. For supervised style transfer, in addition to the source style image, its non-stylized counterpart is also required to learn the transformation between them, as shown in Fig. 1(b). Unfortunately, such a pair of inputs is not readily available in practice, which greatly limits its application scope.

In this work, I handle a more challenging unsupervised stylization problem, only with a binary text-based binary image and an arbitrary style image as in Fig. 1(a). To bridge the distinct visual discrepancies between the binary image and the style image, we extract the main structural imagery of the style image to build a preliminary mapping to the binary image. The mapping is then refined using a structure transfer algorithm, which adds shape characteristics of the source style to the binary shape. In addition to the distribution constraint [8], a saliency constraint is proposed to jointly guide the texture transfer process for the shape legibility. These improvements allow our unsupervised style transfer to yield satisfying artistic results without the ideal input required by supervised methods.

Furthermore, I investigate the combination of stylized shapes (text, symbols, icons) and background images, which is very common in visual design. Specifically, I propose

a) Unsupervised text-based binary image stylization



b) Supervised text stylization

A new context-aware text-based binary image stylization and synthesis framework, where the target binary shape is seamlessly embedded in a background image with a specified style. By “seamless”, I mean the target shape is stylized to share context consistency with the background image without abrupt image boundaries, such as decorating a blue sky with cloud-like typography. To achieve it, we leverage cues considering both seamlessness and aesthetics to determine the image layout, where the target shape is finally synthesized into the background image. When a series of different styles are available, our method can generate diverse artistic typography, symbols or icons against the background image, thereby facilitating a much wider variety of aesthetic interest expression. In summary, our major technical contributions are:

- I raise a new text-based binary image stylization and synthesis problem for visual design and develop the first automatic aesthetic driven framework to solve it.
- I present novel structure and texture transfer algorithms to balance shape legibility with texture consistency, which we show to be effective in style transition between the binary shape and the style image.
- I propose a context-aware layout design method to create professional looking artwork, which determines the image layout and seamlessly synthesizes the artistic shape into the background

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 8, No. 1, August : 2023

image.

The rest of this paper is organized as follows. In Section II, we review related works in style transfer and text editing. Section III defines the text-based binary image stylization problem, and gives an overview of the framework of our method. In Section IV and V, the details of the proposed legibility-preserving style transfer method and context-aware layout design method are presented, respectively. I validate our method by conducting extensive experiments and comparing with state-of-the-art style transfer algorithms.

II. RELATED WORK

Pioneering methods transfer colors by applying a global transformation to the target image to match the color statistics of a source image [10]–[12]. When the target image and the source image have similar content, these methods generate satisfying results. Subsequent methods work on color transfer in a local manner to cope with the images of arbitrary scenes. They infer local color statistics in different regions by image segmentation [13], [14], perceptual color categories [15], [16] or user interaction [17]. More recently, Shih et al. employed fine-grained patch/pixel correspondences to transfer illumination and color styles for landscape images [18] and headshot portraits [19]. Yan et al. [20] leveraged deep neural networks to learn effective color transforms from a large database. In this paper, we employ color transfer technology [11] to reduce the color difference between the style image and the background image for seamless shape embedding. Texture synthesis technologies attempt to generate new textures from a given texture example. Non-parametric methods use pixel [21] or patch [1] samplings in the example to synthesize new textures. For these methods, the coherence of neighboring samples is the research focus, where patch blending via image averaging [22], dynamic programming [1], graph cut [23] and coherence function optimization [24] is proposed. Meanwhile, parametric methods build mathematic models to simulate certain texture statistics of the texture example. Among this kind of methods, the most popular one is the Gram-matrix model proposed by Gatys et al. [25]. Using the correlations between multi-level deep features to represent textures, this model produces natural textures of noticeably high perceptual quality. In this paper, we adapt conventional texture synthesis methods to dealing with binary text images. We apply four constrains of text shape, texture distribution, texture repetitiveness and text saliency to the texture synthesis method of Wexler [24] to build our novel texture transfer model. In texture transfer, textures are synthesized under the structure constraint from an additional content image. According to whether a guidance map is provided, texture transfer can be further categorized into supervised and unsupervised methods. Supervised methods, also known as image analogies [9], rely on the availability of an input image and its stylized result. These methods learn a mapping between such an example pair, and stylize the target image by applying the learned mapping to it. Since first reported in [9], image analogies have been extended in various ways such as video analogies [26] and fast image analogies [27]. The main drawback of image analogies is the strict requirement for the registered example pair. Most often, we only have a style image at hand, and need to turn to the unsupervised texture transfer methods. Without the guidance of the example pair, unsupervised methods directly find mappings between different textures.

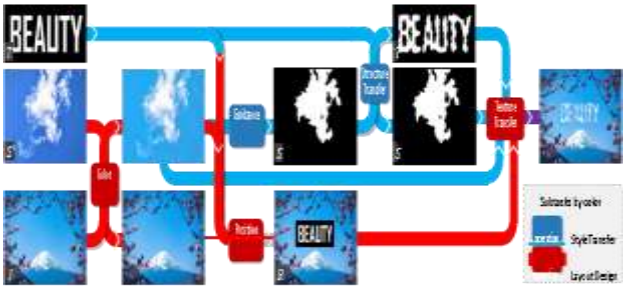


Fig 2 Overview of an algorithm

modalities. For instance, Efros and Freeman [1] introduced a guidance map derived from image intensity to help find correspondences between two texture modalities. Zhang *et al*. [28] used a sparse-based initial sketch estimation [29] to construct a mapping between the source sketch texture and the target image. Frigo *et al*. [3] put forward a patch partition mechanism for an adaptive patch mapping, which balances the preservation of structures and textures. However, these methods attempt to use intensity features to establish a direct mapping between the target image and the style image, and will fail in our case where the two input images have huge visual differences. By contrast, our method proposes to extract an abstract binary imagery from the style image, which shares the same modality as the target image and serves as a bridge. Fueled by the recent development of deep learning, there has been rapid advancement of deep-based methods that leverage high-level image features for style transfer. In pioneering Neural Style [2], the authors adapted Gram-matrix-based texture synthesis [25] to style transfer by incorporating content similarities, which enables the composition of different perceptual information. This method has inspired a new wave of research on video stylization [30], perceptual factor control [31] and acceleration [32]. In parallel, Li and Wand [6] introduced a framework called CNNMRF that exploits Markov Random Field (MRF) to enforce local texture transfer. Based on CNNMRF, Neural Doodle [33] incorporates semantic maps for analogy guidance, which turns semantic maps into artwork. The main advantage of parametric deep-based methods is their ability to establish semantic mappings. For instance, it is reported in [6] that the network can find accurate correspondences between real faces and sketched faces, even if their appearances differ greatly in pixel domain. However, in our problem, the plain text image provides little semantic information, making these parametric methods lose their advantages in comparison to our non-parametric method, as demonstrated in Fig. 12.

In the domain of text image editing, several tasks have been addressed like calligrams [34]–[36] and handwriting generation [37], [38]. Lu et al. [39] arranged and deformed pre-designed patterns along user-specified paths to synthesize decorative strokes. Handwriting style transfer [40] is accom- plished using non-parametric samplings from a stroke library created by trained artists or parametric neural networks to learn stroke styles [38]. However, most of these studies focus on text deformation. Much less has been done with respect to the fantastic text effects such as shadows, outlines, dancing flames (see Fig. 1), and soft clouds (see Fig. 2).

To the best of our knowledge, the work of Yang et al. [8] is the only prior attempt at generating text effects. It solves the text stylization problem using a supervised texture transfer technique: a pair of registered raw text and its counterpart text effects are provided to calculate the distribution characteristics of the text effects, which guide the subsequent texture synthe- sis. In contrast, our framework automatically generates artistic typography, symbols and icons based on arbitrary source style images, without the input requirements as in [8]. Our method provides a more flexible and effective tool to create unique visual design artworks.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 8, No. 1, August : 2023

III. PROBLEM FORMULATIO8N AND FRAMEWORK

I aim to automatically embed the target text-based binary shape in a background image with the style of a source reference image. To achieve this goal, I decompose the task into two subtasks: 1) Style transfer for migrating the style from source images to text-based binary shapes to design artistic shapes. 2) Layout design for seamlessly synthesizing artistic shapes in the background image to create visual design artwork such as posters and magazine covers.

Fig. 2 shows an overview of our algorithm. For the first subtask, I abstract a binary image from the source style image, adjust its contour and the outline of the target shape to narrow the structural difference between them. The adjusted results establish an effective mapping between the target binary image and the source style image. Then Iam able to synthesize textures for the target shape. For the second subtask, I first seek the optimal layout of the target shape in the background image. Once the layout is determined, the shape is seamlessly synthesized into the background image under the constraint of the contextual information. The color statistics of the background image and the style image are optionally adjusted to ensure visual consistency.

IV.TEXT-BASED BINARY IMAGE STYLE TRANSFER

A. Guidance Map Extraction

The perception of texture is a process of acquiring abstract imagery, which enables us to see concrete images from the disordered (such as clouds). This inspires us to follow human's abstraction of the texture information to extract the binary imagery. S serves as a guidance map, where white pixels indicate the reference region for the shape interior (foreground) and black pixels for the shape exterior (background). The boundary of foreground and background depicts the morphological characteristics of the textures in SJ. We propose a simple yet effective two-stage method to abstract the texture into the foreground and the background with the help of texture removal technologies.
In particular, we use the Relative Total Variation (RTV) [42] to remove the color variance inside the texture, and obtain a rough structure abstraction. However, texture contours are also smoothed in (see Fig. 3(b)(f)). Hence, we put forward a two-stage abstraction method. In the first stage, pixels are abstracted as fine-grained super pixels [43] to precisely match the texture contour. Each super pixel uses its mean pixel values as its feature vector to avoid the texture variance. In the second stage, the super pixels are further abstracted as the coarse-grained foreground and background via K-means clustering (K = 2). Fig. 3 shows an example where our two-stage method generates accurate abstract imagery of the plaster wall. In this example, our result has more details at the boundary than the one-stage method, and fewer errors than the state-of-the-art label-map extraction method [41] (see the zoomed region in Fig. 3(h)).

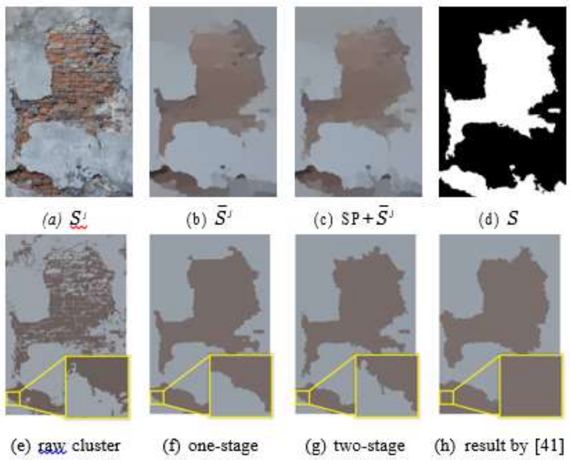


Fig. 3. Guidance map extraction.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 8, No. 1, August : 2023

B. Structure Transfer

Directly using S extracted in Section IV-A and the input T for style transfer results in unnatural texture boundaries as shown in Fig. 4(a). A potential solution could be employing the shape synthesis technique [47] to minimize structural inconsistencies between S and T. In Layered Shape Synthesis (LSS) [47], shapes are represented as a collection of boundary patches at multiple resolution, and the style of a shape is trans- ferred onto another by optimizing a bidirectional similarity function. However, in our application such an approach does not consider the legibility, and the shape will become illegible after adjustment as shown in the second row of Fig. 6. Hence, we incorporate stroke trunk protection mechanism into LSS and propose a legibility-preserving structure transfer method. The main idea is to adjust the shape of the stroke ends while preserving the shape of the stroke trunk, because the legibility of a glyph is mostly determined by the shape of its trunk.

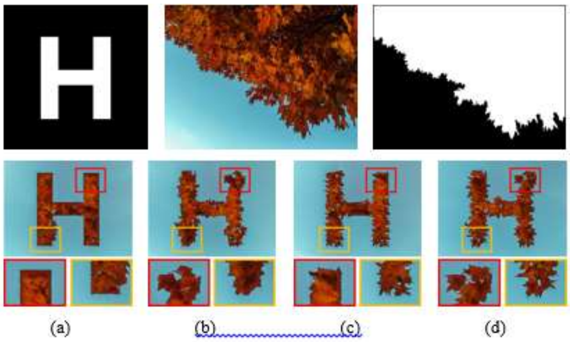


Fig. 4. Benefits of bidirectional structure transfer. The forward transfer simulates the distribution of leaves along the shape boundary, while the backward transfer generates the fine details of each leaf shape. Their combination creates vivid leaf-like typography.

V. CONTEXT-AWARE LAYOUT DESIGN

A. Color Transfer

Obvious color discontinuities may appear for style images SJ that have a different color from the background I. Therefore, we employ color transfer technology. Here we use a linear method introduced by Image Analogies color transfer [11]. This technique estimates a color affine transformation matrix and a bias vector which match the target mean and standard deviation of the color feature with the source ones. In general, color transfer in a local manner is more robust than the global method. Hence, we employ the perception-based color clustering technique [15] to divide pixels into eleven color categories. The linear color transfer is performed within corresponding categories. More sophisticated methods or user interactions could be optionally employed to further improve the color transfer result.

B. Symbol and Icon Rendering

The proposed method has the ability to render textures for text-based geometric shapes such as symbols and icons. Fig. 5 shows that our method successfully transfers rippling textures onto the binary Zodiac symbols. It seems that the proposed method is also capable of stylizing more general shapes, like the emoji icons in Fig. 6. Meanwhile, we notice that our saliency term selects the prominent orange moon to be synthesized into the sun and heart, which enriches the color layering of the results.

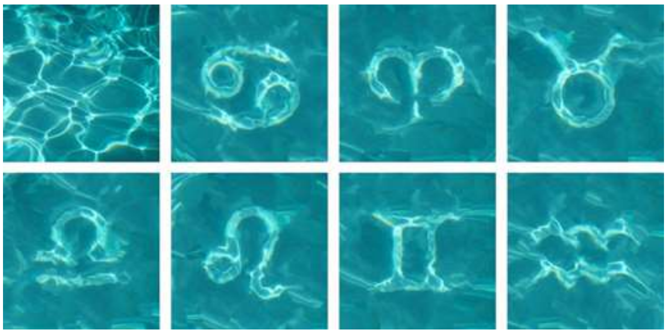


Fig. 5. Rendering rippling Zodiac symbols using a photo of water

Fig. 6. Rendering emoji icons with the painting style of Van Gogh using "*The Starry Night*"

VI. EXPERIMENTAL RESULTS AND ANALYSIS

A. Visual Textual Presentation Synthesis

I aim to synthesize professional looking visual-textual presentation that combines beautiful images and overlaid stylish text. In Fig. 7, two visual-textual presentations automatically generated by our method are provided. In the example barrier reef, a LOVE-shaped barrier reef is created, which is visually consistent with the background photo. I further show in the example cloud that I can integrate completely new elements into the background. Clouds with a specific text shape are synthesized in the clear sky. This approach is capable of artistically embellishing photos with meaningful and expressive text and symbols, thus providing a flexible and effective tool to create original and unique visual-textual presentations.

Fig.7 Visual textual presentation synthesis

B.Running Time

When analyzing the complexity of the proposed method, I consider the time of guidance map extraction, position estimation and color/structure/texture transfer. To simplify the analysis, we assume the target image T has N pixels, and the image resolution have the same magnitude O(N) as T . In addition, the patch size and the number of iterations is constants that can be ignored in computational complexity. Guidance map extraction. According to [42]–[44], the complexity of

RTV, super pixel extraction and saliency detection is O(N). K-means has a practical complexity of O(NKt) [53], where K = 2 and t is the number of iterations. Ignoring K and t, the total complexity of guidance map extraction is O(N). *Style transfer*. According to [11], color transfer has an $O(N)$ complexity. During structure transfer, patches along the shape boundary are matched using FLANN [54]. The upper bound of the patch number is $O(N)$ and thus the proposed structure transfer is $O(N \log N)$ complex. As reported in [48], Patch-

Match in texture transfer has a complexity of $O(N \log N)$.In summary, the overall computational complexity of the proposed method is $O(N \log N)$.

VII. CONCLUSION AND FUTURE WORK

In this paper, I demonstrate a new technique for text- based binary image stylization and synthesis to incorporate binary shape and colorful images. I exploit guidance map extraction to facilitate the structure and texture transfer. Cues for seamlessness and aesthetics are leveraged to determine the image layout. Our context-aware text-based image stylization and synthesis approach break through a barrier between images and shapes, allowing users to create fine artistic shapes and to design professional looking visual-textual presentations. There are still some interesting issues for further investigation. A direction for future work is the automatic style image selection.

REFERENCES

[1] A. A. Efros and W. T. Freeman, “Image quilting for texture synthesis and transfer,” in *Proc. ACM Conf. Computer Graphics and Interactive Techniques*, 2001, pp. 341–346.
[2] L. A. Gatys, A. S. Ecker, and M. Bethge, “Image style transfer using convolutional neural networks,” in *Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition*, 2016.
[3] O. Frigo, N. Sabater, J. Delon, and P. Hellier, “Split and match: example-based adaptive patch sampling for unsupervised style transfer,” in *Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition*, 2016.
[4] M. Elad and P. Milanfar, “Style transfer via texture synthesis,” *IEEE Transactions on Image Processing*, vol. 26, no. 5, pp. 2338–2351, 2017.
[5] J. Liao, Y. Yao, L. Yuan, G. Hua, and S. Kang, “Visual attribute transfer through deep image analogy,” *ACM Transactions on Graphics*, vol. 36, no. 4, p. 120, 2017.
[6] C. Li and M. Wand, “Combining markov random fields and con- volutional neural networks for image synthesis,” in *Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition*, 2016.
[7] D. Chen, L. Yuan, J. Liao, N. Yu, and G. Hua, “Stylebank: An explicit representation for neural image style transfer,” in *Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition*, 2017.
[8] S. Yang, J. Liu, Z. Lian, and Z. Guo, “Awesome typography: Statistics- based text effects transfer,” in *Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition*, 2017.
[9] A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin, “Image analogies,” in *Proc. Conf. Computer Graphics and Interactive Techniques*, 2001, pp. 327–340.
[10] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley, “Color transfer between images,” *IEEE Computer Graphics and Applications*, vol. 21, no. 5, pp. 34–41, 2001.
[11] A. Hertzmann, “Algorithms for rendering in artistic styles,” Ph.D. dissertation, New York University, 2001.
[12] F. Pitié, A. C. Kokaram, and R. Dahyot, “Automated colour grading using colour distribution transfer,” *Computer Vision and Image Under- standing*, vol. 107, no. 1, pp. 123–137, 2007.
[13] Y. W. Tai, J. Jia, and C. K. Tang, “Local color transfer via probabi- listic segmentation by expectation-maximization,” in *Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition*, 2005, pp. 747–754.
[14] “Soft color segmentation and its applications.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 9, pp. 1520– 1537, 2007.