Guorong Xiao[1], **Xuemiao Xu**[2]

GuangDong University of Finance(1), South China University of Technology(2)

An Optimization Method of Text Image

Abstract. *This paper presents a novel method to generate outline-based text image, that approximates the major line structure of the underlying image content, and hence the resultant text fits nicely into the small text screen of mobile devices. We mimic how text image artists deform the underlying image in order to approximate the image content with the limited variety of character font structure. The generation is formulated as an optimization problem that minimizes the shape dissimilarity. Convincing results are shown to demonstrate the effectiveness of the proposed method.*

Streszczenie. *W artykule przedstawiono nową metodę tworzenia obrazu tekstu, na podstawie szkicu. Algorytm aproksymuje strukturę głównej linii zawartej w obrazie w tle, a tekst wynikowy dopasowuje się do małych ekranów, jak w telefonach komórkowych. W metodzie naśladowany jest sposób odkształcenia obrazu w tle, w celu określenia jego zawartości, przy pomocy danej różnorodności znaków. Algorytm pozwala na minimalizację różnic w kształcie obrazów. (**Metoda optymalizacji obrazu tekstu**).*

Keywords: Text image, Shape similarity, Grid deformation
Słowa kluczowe: obraz tekstu, podobieństwo kształtu, deformacja siatki.

Introduction

With the wide usage of text image, its automatic generation is desirable. However, existing methods can only handle the easier tone based text image, as its generation can be regarded as a dithering problem with characters.

In this paper, we formulate the generation process as an optimization problem by minimizing the deformation of the character grid, the number of characters used, and the shape dissimilarity between the character shapes and the underlying image structure. The character grid deformation mimics how text imageists deform the image. Even after the deformation, the shape of characters and the underlying structure can be substantially different. Hence, alignment sensitive metric such as SSIM [9] is not applicable to measure the shape similarity. The translation, scale and rotational invariant shape context [1] [6] is also not applicable as our application requires the selected characters to have a similar position, scale and orientation, as the underlying structure. We propose a novel shape similarity metric that can tolerate the misalignment while remain accounting for position, scale, orientation, and structure. The effectiveness of the method is demonstrated by several examples.

Optimization Method of Text Image

To the best of our knowledge, there is no previous academic study on text image techniques. To produce text image, one can type it with a standard text editor. But it is not as intuitive as painting. To facilitate the painting of text image, people developed interactive painting software to produce text image. Users can directly paint the characters with a mouse or a tablet via a painting metaphor. The only difference is that the output are not pixels, but text characters.

Some research also attempts to automatically convert images to text image. However, they can only generate the tone-based text image as the tone-based one can be regarded as a simple dithering process. Note that the tonebased text image normally consumes more text characters and the result may not fit into the limited text screen. In this paper, we focus on the generation on the outline-based text image as it presents a clearer picture in a smaller text space. Its generation can no longer be regarded as a dithering process. Instead, the shape and structure similarity should play a major role in its generation.

Our framework, given an image (real photograph or cartoon), we start by obtaining its outline for the following generation of outline-based text image. This outline can be simply obtained by naive edge detection. Instead, we employ a more sophisticated line art generation method proposed by Kang et al. [4]. To focus only on the shape of the outline & character images and avoid the influence of line thickness, we perform thinning on the outline & character images so that all lines are with single pixel width. The thinned image is further vectorized, so that it can be rasterized in arbitrary scale.

As the limited shapes of text characters cannot represent all possible image content, artists slightly deform the image in order to allow the combination of characters to represent the deformed image. We mimic this by iteratively deforming a grid overlaid on the outline image. The initial grid is regularly laid. During each iteration, the current grid is deformed and the underlying image is rectified. Each grid cell content is mapped to a rectangular block and approximated by a best-matched character. An objective function is proposed to evaluate the current grid. An optimal grid is selected by minimizing the text resolution, the deformation of text grid, and the dissimilarity between the characters and the rectified image. The dissimilarity is measured according to a novel misalignment-tolerant shape similarity metric. Once the optimization is completed, we obtain the optimal text grid together with the associated text image.

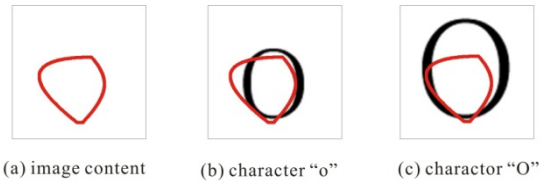


Fig.1. Matching image content with limited character shapes.(a) The image content of an imperfect circle. (b) Character “o” is a reasonable match in terms of shape and position, even it is slightly misaligned with the imperfect circle. (c) In contrast, capital letter “O” is less desirable, as its shape is much different from the image content.

Due to the limited number of characters, it is very likely that the arbitrary image content cannot be well represented by the limited shapes of characters, even with the deformation. Therefore, we need a shape similarity metric that is tolerable to misalignment and, simultaneously, accounts for the image structure. Fig.1.(a) shows an imperfect circle positioned near the bottom. In our application, we can accept representing this imperfect circle with a character “o” even they are not well aligned (Fig.1.(b)). Moreover, we prefer the small letter “o” instead of the capital letter “O” (Fig.1.(c)) as its shape is closer to the underlying content.

Existing shape similarity metrics can be roughly classified into two extreme categories, alignment sensitive metrics and transformation invariant metrics. PSNR (peak signal-to-noise ratio) or MSE (mean squared error), and the well-known SSIM [9] belong to the former category. Their similarity values drop significantly when two equal images are slightly misaligned during the comparison.

On the other hand, the transformation invariant metrics are designed to be invariant to translation, scaling, or orientation. These metrics include shape context descriptor [1] [6], Fourier descriptor, skeleton-based shape matching [3] [7] [8], and curvature-based shape matching [2] [5]. In our application, we need, however, a metric that can tolerate misalignment while remain accounting for position, scale, orientation, and structure.

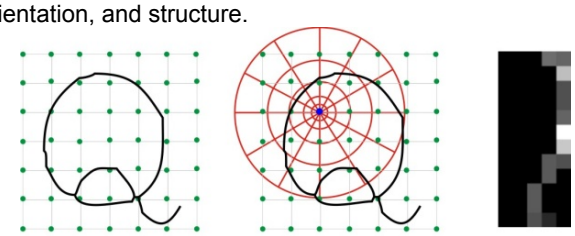


Fig.2. Misalignment-tolerant shape similarity. (a) Regularly sampling the local shape configuration. (b) For each sample, a logpolar histogram is computed. (c) The plot of a feature vector corresponding to the blue reference sample in (b). The row and column correspond to the angular and radial dimensions of the log-polar diagram respectively.

To achieve this, we measure the shape similarity using shape features regularly distributed over the interested region. Given a shape (e.g. Fig.2.(a)) positioned within an interested region, we regularly distribute N sample points. For each sample point, a shape descriptor is employed to quantify the local shape feature.

To tolerate the misalignment, we compute the log-polar histogram[6] as our shape descriptor, since it is insensitive to small shape perturbation. This histogram expresses the local configuration of the shape relative to the reference sample point. As shown in Fig.2.(b), we construct bins that are uniformly distributed in the log-polar space. For each bin, we sum the grayness of the shape and refer the sum as the corresponding component in the shape feature vector. Here the pixel with black color has the grayness of while the one with white color has the grayness of . The bin value $h_i(k)$ of the k-th bin is computed:

$$h_i(k) = \sum_{(q-p_i)\in bin(k)} I(q) \quad (1)$$

q is the position of the current pixel; $q-p_i$ is the relative position to the reference sample point position p_i; I(q) returns the intensity at q. Fig.2.(c) plots the feature vector hi with respect to the sample point pi (blue dot in Fig.2.(b)).

In all of our experiments, we empirically construct histograms with 5 bins on the radial axis in log space, and 12 bins on the angular axis. The radius of the coverage is selected so that the log-polar diagram can roughly cover a character image, that is about half of the side of a character. To reduce the aliasing issue due to the discrete nature of the bins, we blur the image by Gaussian filter of size 7*7 before measuring the shape feature.

The feature vectors of the sample points are then concatenated to describe the shape in the interested region. The shape similarity between two regions S and S ' is measured by comparing their feature vectors, given by

$$D(S,S') = \frac{1}{M}\sum_{i\in N} \| h_i - h_i' \| \quad (2)$$

where hi (hi') is the feature vector of the i-th sample point on S(S '); M is the normalization factor defined as min(n,n '), where (n,n ') is the total grayness of the region S(S '). This normalization factor counteracts the effect of absolute grayness.

The effectiveness of this metric can be demonstrated by Fig.3, in which we query three different shapes on three rows. To evaluate our metric, we compare it to the classical shape context(a translation- and scale- invariant metric), SSIM (an alignmentsensitive, structure similarity metric), and the RMSE (root mean squared error) after blurring. For the last metric, we measure RMSE after blurring the compared images by a Gaussian kernel of 7*7. This RMSE metric is compared because one may argue our metric is similar to RMSE after blurring the images. For each metric, we list the best match and the first runner-up.

From the results, the shape context over-emphasizes on the shape while ignoring the position and scale (especially in query3). On the contrary, SSIM is vulnerable to the misalignment. Almost in all cases, it returns characters that are not similar to the queries. Although the RMSE-after-blurring metric can tolerate misalignment and pay more attention to the position, it fails to account for the shape and structure, almost in all queries. In contrast, our metric returns reasonable results in all queries. It can well balance the structure, orientation, location, scale, and simultaneously, tolerate the misalignment.

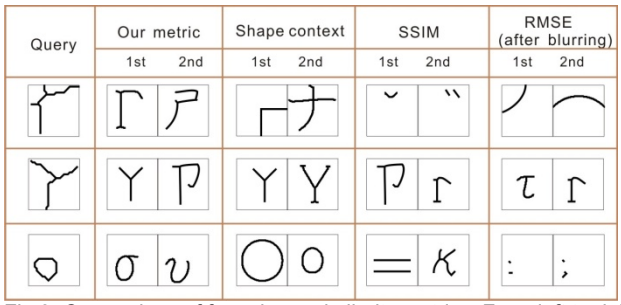


Fig.3. Comparison of four shape similarity metrics. From left to right: our metric, shape context, SSIM and RMSE-after-blurring. Three queries are made on the three row. For each metric, we list the best match and the first runner-up.

To overcome the limited shape variety of the limited characters, artists de*form the underlying image. We mimic such deformation via an optimization process. With the proposed shape similarity metric, we can then determine the optimal character grid resolution as well as how to deform the grid optimally. In this section, we focus on the grid deformation given the grid resolution, as the deformation is core of the optimization.*

We start by overlaying the underlying image with a character grid of a given resolution of k. Each cell contains $L_h * L_v$ pixels, same as the resolution of the character fonts. Note that we only handle fixed-width fonts. With this layout, we can already determine the best matched character for each grid cell using Eq.(2). Of course, there may be matching error between the cell content and the best-matched character. The sum of these errors gives us one factor to quantify the fitness of the current grid layout. Another factor to consider is the degree of deformation. An over-deformed grid cannot give us a good text image even all cells are perfectly matched with characters. Therefore, we define the following objective function to balance between the character matching and the degree of deformation.

Besides the interior grid vertices, we also allow the vertices on the outermost boundary to move, instead of fixing them at the boundary. This is equivalent to allow the whole

grid to translate globally. In our current implementation, we employ the simulated annealing to solve this discrete optimization problem.

Results and Discussions

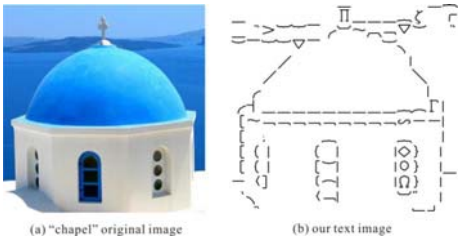
To validate our method, we test our method with a variety of input images, including real photographs and cartoons. Our character database contains 314 distinct characters. They include 96 characters from ASCII table, 55 Greek characters, 90 Japanese Hiragana and Katakana, 64 Russian characters, and 9 table symbols (component lines to form table).

Fig.5 shows the results of converting cartoon to outline-based text image. Note that how the eyes in them are faithfully represented. Fig.4 shows the result of a real photograph. In general, the higher the complexity of the image content, the higher text grid resolution is needed.

Performance and Limitations

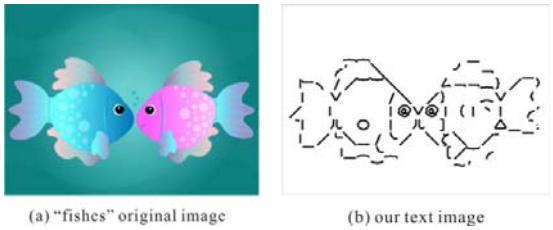
The proposed system is implemented on a PC with CPU 2.00 GHz, 8 GB system memory, and nVidia Geforce GTX 280 GPU with 1G video memory. To achieve high performance, we realize several parts of the system on GPU. The matching, rectifying, thinning, and blurring are all executed in GPU. The data transfer between GPU memory and main memory is also minimized to avoid transfer bottleneck.

Besides the traditional text image that only works on fixed-width font, modern text image deals also with proportional fonts, e.g. Japanese Shift-JIS. Our current method does not handle proportional fonts. To extend our work to support proportional font, the grid layout has to be changed as the number of characters on a horizontal line is data-dependent and not fixed.



(a) "chapel" original image (b) our text image

Fig. 4. "chapel".



(a) "fishes" original image (b) our text image

Fig. 5. Fishes.

Conclusion

In this paper, we formulate the generation process as an optimization problem by minimizing the deformation of the character grid, the number of characters used, and the shape dissimilarity between the character shapes and the underlying image structure. We present a method that mimics how artists deform the given underlying image. Our method is formulated as an optimization to balance between the shape similarity, the text grid deformation, and the text grid resolution. In order to match the shape, a novel misalignment-tolerant metric is proposed that accounts for position, scale, orientation, and structure. Several convincing results, that are comparable to manually prepared text image, have been shown to demonstrate the effectiveness of our method.

Acknowledgements

This work was supported by National Natural Science Foundation of China(Grant No. 61103120).

REFERENCES

[1] Belongie S., Malik J., and Puzicha J.: Shape matching and object recognition using shape contexts, IEEE Transactions on Pattern Analysis and Machine Intelligience, 24 (2002):509-522

[2] Cohen I., Ayache N., and Sulger P.: Tracking points on deformable objects using curvature information. In ECCV'92: Proceedings of the Second European Conference on Computer Vision, Springer-Verlag, (1992):458-466

[3] Goh W.B.: Strategies for shape matching using skeletons. Comput. Vis. Image Underst. 110(2008):326-345

[4] Kang H., Lee S., Chui C.K.: Coherent line drawing. In ACM Symposium on Non-Photorealistic Animation and Rendering (NPAR), (2007):43-50

[5] Milios E.E.: Shape matching using curvature processes. Comput. Vision Graph. Image Process. 47(1989):203-226

[6] Mori G., Belongie S., Malik J.: Efficient shape matching using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(2005):1832-1837

[7] Sundar H., Silver D., Gagvani N., Dickinson S.: Skeleton based shape matching and retrieval. SMI '03:Proceedings of the Shape Modeling International, 2003

[8] Torsello A., Hancock E.R.: A skeletal measure of 2d shape similarity. Computer Vision and Image Understanding, 95(2004):1-29

[9] Wang Z., Bovik A.C., Sheikh H.R., Member S., SI Moncelli E. P., Member S.: Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing, 13(2004): 600-612

Authors: *Lecturer Guorong Xiao, School of Computer Science and Technology , GuangDong University of Finance, 510000, Guangzhou China, Email:newducky@126.com; Associate Professor Xuemiao Xu, School of Computer Science and Engineering, South China University of Technology, 510000, Guangzhou, China,Email: xm_winsy@163.com, Corresponding author.*