SIGNAL RICH ART IMAGE — A NEW TOOL FOR AUTOMATIC IDENTIFICATION AND DATA CAPTURE APPLICATIONS USING MOBILE PHONES

Ya-Lin Lee and Wen-Hsiang Tsai

Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan

ABSTRACT

A new tool, called signal-rich-art image, for automatic identification and data capture applications is proposed, which is created from an artistic target image for use as a carrier of a given message. The created image is visually similar to the target image, achieving the effect of so-called signal rich art. With its function similar to those of barcodes or QR codes, such a type of image is produced by fragmenting the composing characters of the message and injecting them into the target image by a novel image-block luminance modulation scheme. Skillful techniques are also proposed for message extraction from a mobile phone-captured version of the signal-rich-art image printed on paper or displayed on a screen. Good experimental results show the feasibility of the proposed method.

Index Terms—signal rich art, automatic identification and data capture, barcodes, QR codes, signal rich art image.

1. INTRODUCTION

Signal rich art, as defined by Davis [1], is *the art that communicates its identity to context-aware devices*. A type of image of this nature, called *signal-rich-art image*, is proposed in this paper. The image may be printed as a hardcopy for use of any purpose, which is then "re-imaged" by a mobile-phone camera and "understood" by some *automatic identification and data capture* (AIDC) techniques [2] proposed in this study. Printed signal-rich-art images may be of the forms of documents, labels, posters, etc. Also, such images may have the visual appearances of artistic-flavored photos, pictures, paintings, which are more attractive to humans than those produced by conventional AIDC techniques, like barcodes, QR-codes, etc.

With the advance of technology, machines have long been used to read automatically information in the reality for various applications, like optical character recognition (OCR), license plate recognition, supermarket checkout systems, etc. Recently, many more methods have been developed for this purpose, and they are collectively known as AIDC techniques [2]. The processed information is presented in various forms, some being visible (like barcodes or non-transparent watermarks [12]) and others invisible (like watermarks hidden behind images [13]). Such forms of information, often presented artistically, are termed integrally as *signal rich art* in [1] as mentioned previously.

One technique that realizes the use of signal rich art for the AIDC purpose is *barcode reading*. Being attached to objects, barcodes represent machine-readable data by patterns of lines, rectangles, dots, etc. But most types of barcodes, such as PDF417, QR code, and Data Matrix code, just encode information to yield unsightly images with no aesthetics. If a barcode contains not only the encoded information but also has a visual appearance of an art image, the artistic effect of the barcode will be more attractive than those of conventional ones. For this, many methods have been proposed recently to embed information into *image barcodes* using halftone techniques [3]-[7]. These barcodes have the visual appearances of other images and the encoded information can be decoded from their *hardcopy* versions acquired by scanners. That is, the encoded information can survive *print-and-scan "attacks."*

However, if one uses a mobile phone to capture images of *hardcopy image barcodes*, the information might not be decoded successfully since the captured image will suffer from more types of distortions than those acquired by scanning, such as geometric deformation, noise addition, blurring, etc. Also, message carriers other rather than printed papers, such as screens on display devices, cannot be used to encode information since the halftone methods are based on the printing technique. Instead, the method proposed in this study can decode the message which is carried in an image captured from a printed paper or a display screen using a mobile-phone camera, achieving the effect of signal rich art.

Subsequently, the idea of the proposed method is described in Sec. 2 and the details are presented in Sec. 3. In Sec. 4, some experimental results are presented, followed by conclusions in Sec. 5.

2. IDEA OF PROPOSED METHOD

The proposed signal-rich-art image not only includes the content of a given message, but also has an artistic effect of being visually similar to the pre-selected target image. The proposed method of using the signal-rich-art image for AIDC purposes is illustrated in Fig. 1, which includes two phases: image generation and message extraction.

In the first phase, given a target image I_t and a message M, a signal-rich-art image I_s is created by three steps: (1) transform M into a *message image* I_m consisting of the characters of the message content; (2) *modulate* the gray

values of each *character-fragment* F_i of I_m into two values calculated from the Y-channel values of the corresponding target block B_i of I_t, resulting in a *modulated message image* I_m'; (3) replace the Y-channel of I_t with I_m' to get the desired I_s.

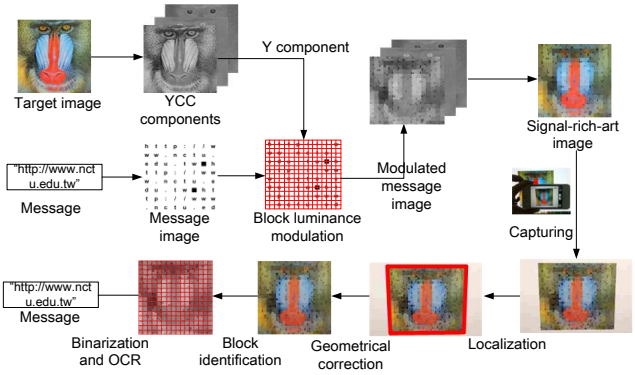


Fig. 1. Illustration of proposed method.

In the second phase, the message M is extracted from a captured version I_s' of a paper or display copy of the signal-rich-art image I_s by four steps: (1) localize and segment out the region I_s'' of the original part of I_s in I_s'; (2) perform an inverse perspective transform to correct the geometric distortion in I_s''; (3) identify the blocks in I_s'', binarize them, and perform OCR to extract the message M from them.

3. SIGNAL-RICH-ART IMAGE GENERATION AND MESSAGE EXTRACTION

3.1. Message image creation

Unlike most barcode systems that encode message contents by patterns (dots, lines, etc.), the proposed method converts a message M into a set of *binary character shapes* drawn from a database, as illustrated in Fig. 2(a). Next, a message image I_m of the size of the target image I_t is created by aligning the character shapes plus an *ending pattern* as shown in Fig. 2(b) in a raster-scan order. If the result cannot fill up I_m, then repetitions of the character shapes are conducted. For example, with the target image I_t as shown in Fig. 2(c) and the message M = "http://www.nctu.edu.tw," the resulting message image I_m is as that shown in Fig. 2(d).

3.2. Block luminance modulation

After the message image I_m is created, it is "injected" into the target image I_t under the constraint that the resulting image retains the visual appearance of I_t. For this, we utilize the characteristic of the YCbCr color model, where the luminance component Y is independent of the others [8], to inject I_m into the Y-channel of I_t. This will solve a problem of illumination variation encountered in later message extraction. A block luminance modulation technique is proposed here to divide the message image I_m into character-fragments F_i and modulate the *mean* of each F_i to be the same as that of the corresponding target block B_i of I_t. The resulting modulated message image I_m' looks like the Y-component of the target image I_t. For example, Fig. 2(f) shows the created modulated message image I_m' which looks like the Y-component of the target image shown in Fig. 2(e), and Fig. 2(g) shows a zoom-out of part of Fig. 2(f) (the red portion).

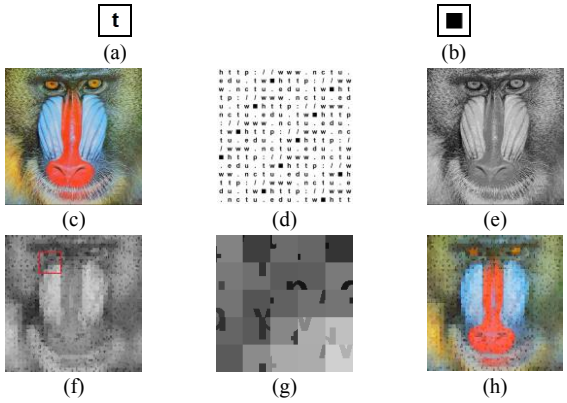


Fig. 2. Signal-rich-art image generation. (a) Image of character "t." (b) Ending pattern. (c) Target image. (d) Message image. (e) Y-channel of (c). (f) Modulated message image. (g) Zoom-out of red square region in (f). (h) Resulting signal-rich-art image.

In more detail, firstly the message image I_m and the Y-component of I_t are divided into character-fragments and target blocks, respectively, where the size of each block is 1/4 times of a character image. The character-fragments F_i then are fitted into the target blocks B_i in a random way controlled by a key K.

Secondly, let N_B and N_W denote the numbers of black and white pixels of the character and non-character parts in F_i, respectively. The pixels of each B_i are sorted according to their Y values in an ascending order to obtain an ordered Y-value set $\{q_1', q_2', \ldots, q_n'\}$. Then, two representative Y values r_1 and r_2 are computed for B_i as:

$$r_1 = \frac{1}{N_B}\sum_{i=1}^{N_B} q_i', \quad r_2 = \frac{1}{N_W}\sum_{i=N_B+1}^{N_B+N_W} q_i' \quad . \tag{1}$$

Finally, the value p_i of each pixel P_i in F_i is modulated to obtain a new pixel value p_i' by the following rule:

$$\textit{set } p_i' = r_1 \textit{ if } P_i \textit{ is black; or set } p_i' = r_2 \textit{ if } P_i \textit{ is white.} \tag{2}$$

The mean of F_i' may be verified easily to be identical to that of B_i, meaning that the overall gray appearances of the modulated message image I_m' and the Y-component of I_t are roughly the same. Then, the Y-component of I_t is replaced with I_m' to generate finally the desired signal-rich-art image I_s which has the visual color appearance of I_t, as shown by the example seen in Fig. 2(h).

Accordingly, later while conducting message extraction, the message characters can be extracted from the Y-component of a captured version of I_s by classifying the pixels of each block into two groups according to their Y values, with the two pixel groups representing the character and non-character parts, respectively. Also, an OCR technique is applied to extract these characters. However, if

the two representative values r_1 and r_2 are too close, it is hard to separate them in the message extraction phase. Therefore, the difference between r_1 and r_2 is adjusted to be larger than a certain pre-defined threshold value δ. For example, Fig. 3 shows four character-fragments resulting from modulations with different values of δ.

(a) (b) (c) (d) (e) (f)

Fig. 3. Modulated character-fragments resulting from uses of different threshold values of δ for the difference between the two representative values r_1 and r_2. (a) $\delta = 0$. (b) $\delta = 10$. (c) $\delta = 20$. (d) $\delta = 30$. (e) $\delta = 40$. (f) $\delta = 50$.

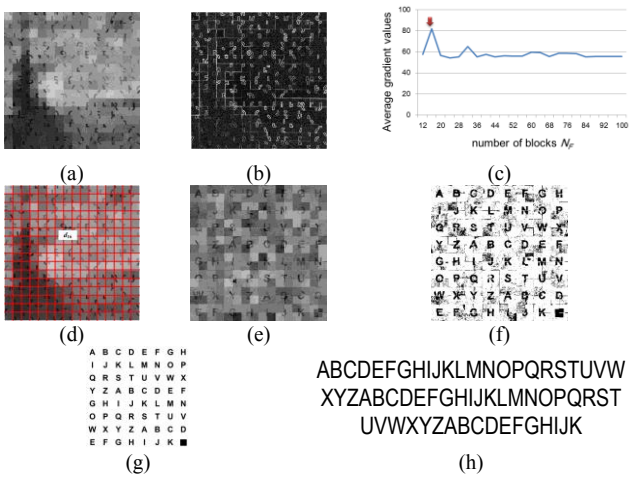
3.3. Signal-rich-art image localization and inverse perspective transform

Assume that the signal-rich-art image I_s is printed and then posted or displayed against a white background, and that the captured image I_d contains only the original image of I_s and the background. To extract the message from I_d, we must *localize* the region of I_s in I_d. For this, we apply the Hough transform and polygonal approximation to find the largest non-white quadrangle Q in I_d. Also, because during image acquisition I_d will suffer from perspective distortion if the axis of the mobile camera is not directed perpendicularly toward the plane of the signal-rich-art image I_s [9], an *inverse perspective transform* is performed on Q to correct the perspective distortion. Finally, the Y-component of the resulting I_d is taken out as an intermediate result, which we call a *captured* modulated message image and denote by I_m''.

3.4. Block number identification and block segmentation

To identify the blocks in I_m'' in order to binarize them and perform OCR to the contents of them, an idea similar to the Hough transform [10] is adopted: use the *pixels' gradient values* to guess the number N_F of blocks in the horizontal or vertical direction in I_m'' because pixels on the splitting lines of the blocks usually have larger gradient values.

In more detail, at first the gradient value of each pixel in I_m'' is computed by a Sobel operator. Next, for each *possible* value n_j of N_F, the distance d_j between the splitting lines of every two adjacent blocks is computed as $d_j = L/n_j$ where L is the side length of the square-shaped I_m''. Then, the horizontal or vertical lines separated by the distance of d_j are taken as *candidate* splitting lines. Also, the *average gradient value* of the pixels on each candidate spitting line is computed. Finally, the value n_j of N_F with the *largest* average gradient value is taken as the desired number of blocks of I_m'' in the horizontal or vertical direction, and division of I_m'' into blocks is conducted accordingly. For example, Fig. 4(a) shows a captured modulated message image I_m'', Fig. 4(b) is the image of the computed gradient values, and Fig. 4(c) illustrates the average gradient values for different N_F, where the n_j with the largest average gradient value is seen to be 16 (indicated by the red arrow). Therefore, the found value for N_F is 16. The corresponding image division result is shown in Fig. 4(d).



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 4. Message extraction. (a) Captured modulated message image I_m''. (b) Gradient values of (a). (c) Average gradient values of pixels on candidate spitting lines for different N_F. (d) Image division result according to determined number of blocks N_F = 16. (e) Fragment reordering result of (d). (f) Binarization result of (e). (g) OCR result of (f). (h) Extracted message.

3.5. Binarization and optical character recognition

After the blocks of I_m'' are segmented out, the character-fragments of the message image I_m may be recovered from the blocks by using the key K mentioned previously. Denote the resulting image by I_m'''. Then, moment-preserving thresholding [11] is applied to F_i' to binarize it, and every four mutually-connected binarized blocks are grouped up to form a character image I_{c_i}' since a character image is divided into four blocks in the message image generation phase.

Next, a similar degree sd_{ij} between each I_{c_i}' and each reference character image I_{c_j} in the database is computed:

$$sd_{ij} = 1 - \left(\sum_{x=1}^{m} \left| p_{ix}' - p_{jx} \right|\right) / m , \qquad (3)$$

where the pixels of I_{c_i}' and I_{c_j} are assumed to be $\{p_{i1}', p_{i2}', \ldots, p_{im}'\}$ and $\{p_{i1}, p_{i2}, \ldots, p_{im}\}$, respectively. Finally, each I_{c_j}' is recognized using an OCR scheme according to the computed similarity degree values: the character with the largest similarity is taken as the message delivered by I_{c_j}'. For example, the recovered original message image I_m with its character-fragments reordered using a key is shown in Fig. 4(e), which, after being binarized, results in Fig. 4(f). The OCR result of Fig. 4(f) is shown in Fig. 4(g), and the final extracted message characters are shown in Fig. 4(h).

4. EXPERIMENTAL RESULTS

The proposed system was developed using Microsoft C#.NET and the generated signal-rich-art images were captured with an iPhone 4S. The character image database includes the printable characters of the ASCII codes. A series of experiments using different parameters have been conducted and corresponding statistics plotted (in Fig. 6) to show the accuracy rates of extracted message characters,

including: (1) the threshold δ of the minimum difference between r_1 and r_2; and (2) the number of blocks N_F in the horizontal or vertical direction of the message image.

Figs. 5(a) through 5(c) show three test target images used in the experiments. The corresponding signal-rich-art images generated with parameters N_F = 32 and δ = 40 are shown in Figs. 5(d) through 5(f). These images were all printed to be of the same size of 127×127 mm.

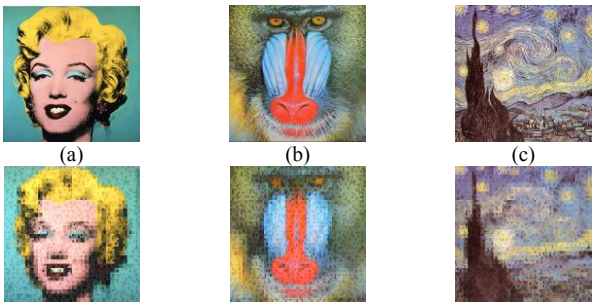


Fig. 5. Created signal-rich-art images. (a)-(c) Test target images. (d)-(f) Resulting signal-rich-art images with N_F = 32 and δ = 40.

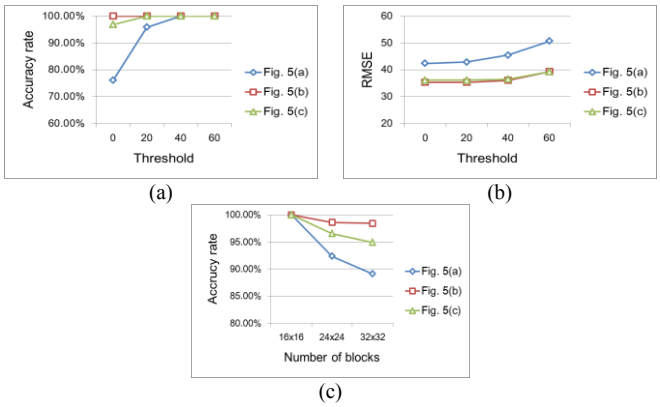


Fig. 6. Plots of trends of results using various parameters. (a) Accuracy rates of extracted messages with different thresholds δ, with #blocks N_F = 16. (b) RMSE values of created signal-rich-art images with respect to target images for different thresholds of δ, with #blocks N_F = 16. (c) Accuracies of extracted messages with different #blocks N_F, where threshold value δ = 40.

One of the parameters that influence the accuracy of the extracted message is the threshold δ for the minimum difference between r_1 and r_2. If δ is too small, the two representative values r_1 and r_2 will be too close so that the message extracted might be wrong. For example, Fig. 6(a) shows the accuracy rate of the extracted messages with δ = 0, 20, 40, and 60, from which it can be seen that the larger the value of δ, the higher the accuracy rate of the extracted message; when δ is larger than 40, an accuracy rate of 100% is yielded.

It can also be seen from Fig. 6(b) that the larger the value of δ, the larger the RMSE of the resulting signal-rich-art image with respect to the target image. So there is a tradeoff between achieving higher message extraction accuracy and obtaining a better visual quality in the resulting signal-rich-art image.

Another parameter that influences the accuracy of the extracted message is the number N_F of blocks in the horizontal or vertical direction in the message image. The larger the value of N_F, the larger the character capacity of the message image; however, the larger the value of N_F, the smaller the size of the block, and so the lower the accuracy of the extracted message, as can be seen from Fig. 6(c).

Finally, to show the robustness of the proposed method, we have conducted some attacks on the created signal-rich-art images. For example, Figs. 7(a) and 7(b) show two attacked versions of the signal-rich-art image shown in Fig. 5(f) with message "ComputerVisionLab" injected. The experimental results show that the carried message can still be extracted from either of these two attacked images. In addition, by regarding image taking from a display screen as a type of attack, a third attacked version so acquired is shown in Fig. 7(c). The resulting message extraction rate is 96.88%, which, though not 100%, means that the proposed method can handle message carriers other than paper copies.

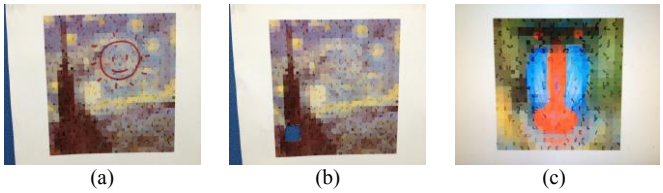


Fig. 7. Robustness of proposed method. (a) A captured signal-rich-art image under defacement attack. (b) A captured signal-rich-art image under another defacement attack. (c) A signal-rich-art image captured from a monitor screen.

5. CONCLUSION

A new type of image for AIDC applications, called signal-rich-art image, has been proposed, which is created from a target image for use as a carrier of a given message. The artistic favor of the target image is kept in the created image, achieving the signal-rich-art effect. Comparing with other AIDC tools like QR codes and hardcopy image barcodes, the proposed signal-rich-art image has several merits: (1) the image can not only be printed on papers but also be displayed on screens for various uses; (2) the image can endure more distortions like perspective transformation, noise, screen blurring, etc.; (3) the message can be extracted from an image captured by a mobile phone (this is not the case for the hardcopy image barcode [3]-[7]); (4) by utilizing the power of OCR, the image can endure more serious attacks, such as partial defacement, image taking from screens, etc. (again, this is not the case for the hardcopy image barcode); (5) if message extraction from the message image by machine is not necessary to carry out, humans can still read the information appearing in the extracted message image because it is composed of characters, and so meaningful and readable. Experimental results show the feasibility of the proposed method.

6. REFERENCES

[1] B. Davis, "Signal rich art: enabling the vision of ubiquitous computing," in *Proc. SPIE* 7880: *Media Watermarking, Security, and Forensics III*, N. D. Memon, J. Dittmann, A. M. Alattar, and E. J. Delp III, Eds., vol. 788002, Feb. 2011.

[2] A. Furness, "Machine-readable data carriers - a brief introduction to automatic identification and data capture," *Assembly Automation*, vol. 20, pp. 28-34, 2000.

[3] D. L. Hecht, "Printed Embedded Data Graphical User Interfaces," *IEEE Computer*, vol. 34, no. 3, pp. 47-55, 2001.

[4] O. Bulan, G. Sharma, and V. Monga, "Orientation modulation for data hiding in clustered-dot halftone prints," *IEEE Trans. Image Processing*, vol. 19, no. 8, pp. 2070-2084, 2010.

[5] O. Bulan, B. Oztan, and G. Sharma, "High capacity image barcodes using color separability," *Proc. SPIE: Color Imaging XVI: Displaying, Processing, Hardcopy, and Applications*, R. Eschbach, G. G. Marcu, and A. Rizzi, Eds., vol. 7866, pp. 20117866-22,1-9, 2011.

[6] N. Damera-Venkata, J. Yen, V. Monga, and B. L. Evans, "Hardcopy image barcodes via block-error diffusion," *IEEE Trans. Image Processing*, vol. 14, no. 12, pp. 1977-1989, 2005.

[7] B. Oztan and G. Sharma, "Continuous phase modulated halftones and their application to halftone data embedding," *Proc. IEEE Intl. Conf. Acoustics Speech and Sig. Proc.*, vol 2, pp. 333-336, 2006.

[8] C. Lin, "Face detection in complicated backgrounds and different illumination conditions by using YCbCr color space and neural network," *Pattern Recognition Letters*, vol. 28, pp. 2190-2200, 2007.

[9] H. Yang, A. C. Kot, and X. Jiang, "Accurate localization of four extreme corners for barcode images captured by mobile phones," *Proc. IEEE Int. Conf. on Image Processing*, pp. 3897-3900, 2010.

[10] J. Illingworth and J. Kittler, "A survey of the Hough transform," *Computer Vision, Graphics, and Image Processing*, vol. 44, no. 1, pp. 87-116, 1998.

[11] W. H. Tsai, "Moment-preserving thresholding: a new approach," *Computer Vision, Graphics, and Image Processing*, vol. 29, no. 3, pp. 377-393, 1985.

[12] T. Y. Liu and W. H. Tsai, "Generic lossless visible watermarking - a new approach," *IEEE Trans. Image Processing*, vol. 19, no. 5, pp. 1224-1235, 2010.

[13] J. Tian, "Reversible data embedding using a difference expansion," *IEEE Trans. Circuits Syst. & Video Technol.*, vol. 13, no. 8, pp. 890-896, Aug. 2003.