COLORIZATION-BASED CODING BY FOCUSING ON CHARACTERISTICS OF COLORIZATION BASES

Shunsuke Ono, Takamichi Miyata, and Yoshinori Sakai

Tokyo Institute of Technology
Oookayama 2-12-1
Meguro-ku, Tokyo, Japan

ABSTRACT
Colorization is a method that adds color components to a grayscale image using only a few representative pixels provided by the user. A novel approach to image compression called colorization-based coding has recently been proposed. It automatically extracts representative pixels from an original color image at an encoder and restores a full color image by using colorization at a decoder. However, previous studies on colorization-based coding extract redundant representative pixels and do not extract the pixels required for suppressing coding error. This paper focuses on the colorization basis that restricts the decoded color components. From this viewpoint, we propose a new colorization-based coding method. Experimental results revealed that our method can drastically suppress the information amount (number of representative pixels) compared with conventional colorization-based coding while objective quality is maintained.

Index Terms— Colorization, Colorization-based coding, Representative pixels, Redundancy, Correct color

1. INTRODUCTION
In recent years, several methods called colorization have been proposed [1][2][3] for adding color to a given grayscale image from a few pixels that have color information. We denote these pixels as representative pixels (RP), and RP can be represented by the positions and color values of these pixels.

Since the information amount for representing positions and color values of RP is small, a novel approach to image compression by using colorization (called colorization-based coding) has been researched [4][5][6]. An encoder extracts RP from an original color image and transmits RP and all luminance components (compressed by the conventional encoder) to a decoder. Then, the decoder restores a color image by colorization.

Obviously, to implement colorization-based coding, automatic RP extraction is required, and which extraction method is chosen determines the performance of the colorization-based coding method. Generally, if we have several sets of RP achieving the same decode quality, the best set is that containing the smallest number of pixels. This means that the RP do not include redundant pixels that contribute little to the quality of the decoded color components (we refer to these pixels as redundant RP).

Cheng et al. [4] and He et al. [5] proposed colorization-based coding that extracts RP based on a machine learning approach. They also discussed the coding efficiencies of their methods by using an objective quality metric such as peak signal-to-noise ratio (PSNR). Cheng et al.’s method adds new pixels to the RP by iterative selection starting from randomly selected initial RP. However, if the initial RP already have some redundancy, there is no procedure for reducing it. In contrast, He et al.’s method selects candidate RP (which could be extracted as RP) as the first step. Then, RP are extracted from the candidate pixels by sequential selection, which guarantees optimality of the machine learning knowledge. However, if the candidate pixels do not initially include the pixels required for suppressing coding error, such pixels can not be extracted at any stage by this method. Furthermore, He et al.’s method does not use the color components of the original image (denoted as original color components) to extract RP. Because of this, their method may not extract the required pixels for RP in many cases. In contrast, the colorization-based coding method proposed by Komiyama et al. [6] extracts RP as a set of color line segments. By restricting the RP to a set of color line segments, the information amount for representing RP is reduced drastically while subjective quality is maintained. However, they did not evaluate their method with any objective quality metric.

In this paper, we present a new colorization-based coding method. Our method first simply segments the original image into squares and extracts RP from each square as initial RP. Then, it reduces the redundant RP in the initial RP by focusing on characteristics of bases of decoded color components which is determined by the positions of the RP and luminance components. We refer to this basis as the colorization bases. Moreover, by using the determined colorization basis and original color components, our method extracts required pixels for RP from pixels that were not extracted as initial RP. In addition, we found that our method can achieve
higher compression than the conventional methods from an objective quality viewpoint.

2. RELATED WORKS

In this section, we describe Levin et al’s colorization algorithm [1], which plays the role of decoder for color components in our proposed colorization-based coding method. We also briefly introduce Cheng et al’s colorization-based coding algorithm [4].

2.1. Colorization

Levin et al’s colorization algorithm is based on a simple premise: neighboring pixels that have similar intensities should have similar colors. We applied their algorithm in the YCbCr color space. Y is the luminance component corresponding to $y$, and Cb or Cr is the color component corresponding to $u$.

Let $n$ be the number of pixels in the original image and $r$ be an identifier of the pixels in raster-scan order ($1 \leq r \leq n$), $u$ ($u \in R^n$) is assumed to be a one-dimensional vector that contains a color component restored by colorization (denoted as the restoration color component) and is arranged in column in raster-scan order. $x$ ($x \in R^n$) is assumed to be a one-dimensional vector that contains RP values, and $x$ has non-zero values only for RP. $u(r)$ and $x(r)$ are the $r$-th elements of $u$ and $x$ respectively. $\Omega = \{r|x(r) \neq 0\}$ is a set of positions of RP. Obviously, $|\Omega|$ is the number of RP that have a specific color value, and it corresponds to the amount of information in-colorization based coding. Let $y(r)$ be a luminance component at the $r$-th pixel. $s \in N(r)$ denotes that the $s$-th pixel is belonging to the neighbor (defined as 8 surrounding pixels) of the $r$-th pixel. Levin et al defined a cost function as

$$J(u) = \sum_{r \in \Omega} \left( u(r) - \sum_{s \in N(r)} w_{rs} u(s) \right)^2 + \sum_{r \in \Omega} (u(r) - x(r))^2,$$  

where $w_{rs}$ is a weighting function that sums to one. Suppose $W$ is an $n \times n$ matrix that contains $w_{rs}$, which is defined as

$$w_{rs} = \begin{cases} 0 & \text{if } r \in \Omega \\ w_{rs} & \text{otherwise}. \end{cases}$$

When $A = I - W$ ($I$ is the $n \times n$ identity matrix) is an affinity matrix, formula (1) is equal to

$$J(u) = ||x - Au||^2.$$  

2.2. Colorization-based Coding by Cheng et al

Cheng et al’s colorization-based coding uses an active learning approach to extract RP automatically. Their method perform better than JPEG for color components. The steps of their method are given below.

1. Divide original image into clusters by image segmentation algorithm.
2. Extract RP randomly from each cluster.
3. Conduct colorization by using temporary RP.
4. Search for clusters that have high error between original and colorized images.
5. Extract more RP from high-error clusters.

Additionally, Cheng et al apply some extension to Levin’s colorization to suit their approach. However, as mentioned in Section 1, their colorization-based coding can not reduce the redundant RP if the initial RP (extracted at step 2) already have redundancy.

3. PROPOSED METHOD

In this section, we present a new colorization-based coding method. In previous methods, there is a high possibility of extracting redundant RP when setting the initial RP. Our approach reduces the redundancy of the initial RP. However, if the initial RP do not include pixels required for suppressing coding error, such beneficial pixels can not be extracted by only the redundancy reduction process. This problem is the same as the problem of He’s method (mentioned in section 1). Thus, we additionally extract required pixels for RP from pixels that were not extracted as initial RP.

The procedure of our proposed method follows.

1. Setting of initial RP.
2. Redundancy reduction of RP.
3. Extraction of required pixels for RP.

3.1. Setting of Initial RP

An original image is divided into squares (the number of squares denoted as $Sq$). We denote a set of pixel positions in raster-scan order that belong to the $k$-th square as $S(k)$ ($1 \leq k \leq Sq$). Then we pick the center pixel of each square as the initial RP.
3.2. Redundancy Reduction of RP

Since the number of RP is very small compared with the number of all pixels, we can say that $x$ is sparse because $x$ has non-zero values only for RP (mentioned in 2.1). Thus, very few columns of $A^{-1}$, which correspond to non-zero elements in $x$, only applied in equation (5). Let $A^{-1}_\Omega$ be the $n \times |\Omega|$ matrix that is composed of columns of $A^{-1}$ that correspond to elements in $\Omega$. Let $x_\Omega$ consist of non-zero elements of $x$, and let $x_\Omega(p) (1 \leq p \leq |\Omega|)$ be the $p$-th element of $x_\Omega$. We can interpret that $u$ is expressed by the sum of columns of $A^{-1}_\Omega$ that treat $x_\Omega$ as a coefficient. Formula (5) is then equal to the following:

$$ u = A^{-1}_\Omega x_\Omega. $$

That is, $u$ is included in a subspace that is spanned by every column vector included by $A^{-1}_\Omega$. These vectors are the colorization bases (mentioned in section 1). Therefore, if there are dependant colorization bases, the RP that correspond to them are redundant. Thus, we can conclude that $u$ changes little even if such RP are deleted.

In our method, using a cosine as a similarity between any two colorization bases, redundancy reduction of RP is achieved as follows:

$$\begin{align*}
\text{set} & \quad \Omega = \Omega / r \quad (x(r) = x_\Omega(j)), \\
\text{if} & \quad \| \langle a_i, a_j \rangle \| / \| a_i \| \| a_j \| \geq Th1,
\end{align*}$$

where $a_i$ and $a_j \ (1 \leq i, j \leq |\Omega|)$ are colorization bases, and $Th1$ is a threshold that is adjusted manually.

3.3. Extraction of Required Pixels for RP

Conventional methods [4][5] determine the values of RP as the original color components corresponding to the positions of the RP, i.e., $x_\Omega$. However, for fixed positions of RP (determined colorization bases), we can compute values of RP referred to as $x'_\Omega$, which minimizes the sum squared error between the decoded and original color components. Note that $x'_\Omega \neq x_\Omega$ in a general case. Let $c$ be a one-dimensional vector that contains an original color component. $x'_\Omega$ is obtained by the following minimization:

$$ x'_\Omega = \arg \min_{\tilde{x}_\Omega} \| A_\Omega \tilde{x}_\Omega - c \|_2. $$

From the results of preliminary experiments, we can observe that the elements in $x_\Omega$ differ greatly from the corresponding elements in $x'_\Omega$ in the areas that need additional RP to decode the color components properly. This fact is easy to understand because, if we have an insufficient number of RP for an area with various colors, the values of the RP must have quite different color components from the original one for minimizing the decoding error in that area. Therefore, we can determine which area needs more RP by comparing the elements of $x_\Omega$ with the corresponding elements in $x'_\Omega$. Then, additional RP are chosen from the local area that have a high difference between $x_\Omega$ and $x'_\Omega$.

In our method, by using the square error between each element of $x_\Omega$ and each element of $x'_\Omega$, extraction of the required pixels for RP is achieved as follows:

$$\begin{align*}
\text{set} & \quad \Omega = \Omega \cup t \quad (r, t \in S(k), x_\Omega(p) = x(r)), \\
\text{if} & \quad \| x_\Omega(p) - x'_\Omega(p) \|_2 \geq Th2,
\end{align*}$$

where $x'_\Omega(p)$ is the $p$-th element of $x'_\Omega$, and $Th2$ is the threshold adjusted manually. Note that $t$ is selected randomly from $S(k)$ ($t \neq r$).

We then obtain the final RP as $x'_\Omega$ corresponding to $\Omega$, which is obtained by (9).

4. EXPERIMENTAL RESULTS

We did an objective test comparing our method with JPEG to evaluate the efficiency of our method. Two color images, called Parrots (256 × 256 [pixel]) and Pepper (256 × 256 [pixel]), shown in Fig. (1) and Fig. (2), respectively were used as original images. We use SSIM (Structural SIMilarity) value as an objective evaluation of image quality. SSIM is the image quality assessment based on the degradation of structural information, and known that it is more closely to human visual estimation than traditional image quality assessment such as PSNR. For complete information about SSIM, see [7]. To obtain the almost same SSIM (for chrominance Cb, Cr) compared with the SSIM of JPEG encoded images, we choose the parameters of our proposed method as follows:

- Parrots : $S_q = 200$, $Th_1 = 0.30$, $Th_2 = 6000$.
- Pepper : $S_q = 200$, $Th_1 = 0.30$, $Th_2 = 5000$.

Table 1. and 2. show the comparisons between the size of color components decoded by our proposed method and those of JPEG. The original color components and decoded results are shown in Figs. 1–6 to visualize the color components effectively, we made the luminance component $Y$ constant for every pixels. For the Parrots image, to achieve the SSIM fixed to 0.8379 for Cb and 0.8474 for Cr, JPEG encoding requires 624 [KByte]. In contrast, our method only needs 472 [KByte] to achieve almost the same SSIM. Similarly, for the Pepper image, our method only needs 528 [KByte] in spite of the fact that the JPEG needs 947 [KByte] for achieving almost the same SSIM. Furthermore, decoded images of our method have less visual artifacts compared with JPEG results. One can see that the decoded images of our method are consisted from only edges and smooth regions.

Besides, we also compare our method with Cheng et al’s method (mentioned in 2.2). They used two color images called Bees (640 × 853 [pixel]) and Girl (512 × 683 [pixel]),
We adjusted the SSIM of the color components same as decoded images by Cheng et al.’s method. As a result, we found that our method achieved almost the same SSIM with quarter size comparing to Cheng et al’s method.

5. CONCLUSION

A new method of colorization-based coding was presented. By focusing on the characteristics of the colorization basis, our method can extract a minimal amount of RP and determine their values. Through this study, the numerical criterion of redundant RP and the relationship between the area lacking RP and the temporal optimal value of RP are clarified.

In the future, we will investigate the automatic setting of parameters and a method of extracting RP independent of the initial RP.

Table 1. Comparison for Parrots image.

<table>
<thead>
<tr>
<th>Method</th>
<th>Size [KByte]</th>
<th>SSIM-Cb</th>
<th>SSIM-Cr</th>
</tr>
</thead>
<tbody>
<tr>
<td>JPEG</td>
<td>624</td>
<td>0.8379</td>
<td>0.8474</td>
</tr>
<tr>
<td>Ours</td>
<td>472</td>
<td>0.8230</td>
<td>0.8552</td>
</tr>
</tbody>
</table>

Table 2. Comparison for Pepper image.

<table>
<thead>
<tr>
<th>Method</th>
<th>Size [KByte]</th>
<th>SSIM-Cb</th>
<th>SSIM-Cr</th>
</tr>
</thead>
<tbody>
<tr>
<td>JPEG</td>
<td>947</td>
<td>0.7504</td>
<td>0.7354</td>
</tr>
<tr>
<td>Ours</td>
<td>528</td>
<td>0.7530</td>
<td>0.7432</td>
</tr>
</tbody>
</table>

6. REFERENCES