Unsupervised regions of interest extraction for color image compression

Xiaoguang Shao (邵晓光)[1,2], Kun Gao (高 昆)[1,2*], Lili Lü (吕丽丽)[1,2], and Guoqiang Ni (倪国强)[1,2]

[1]*School of Optoelectronics, Beijing Institute of Technology, Beijing 100081, China*

[2]*Key Laboratory of Photoelectronic Imaging Technology and System (Beijing Institute of Technology), Ministry of Education, Beijing 100081, China*

[*]*Corresponding author: gaokun@bit.edu.cn*

Received April 12, 2011; accepted June 3, 2011; posted online August 5, 2011

A novel unsupervised approach for regions of interest (ROI) extraction that combines the modified visual attention model and clustering analysis method is proposed. Then the non-uniform color image compression algorithm is followed to compress ROI and other regions with different compression ratios through the JPEG image compression algorithm. The reconstruction algorithm of the compressed image is similar to that of the JPEG algorithm. Experimental results show that the proposed method has better performance in terms of compression ratio and fidelity when comparing with other traditional approaches.

OCIS codes: 100.0100, 100.2000, 150.0150, 150.1135.

doi: 10.3788/COL201210.011001.

At present, lossy compression of still images is one of the research focuses in the digital image-processing field. Under the condition of the limited storage capacity of the imaging system and the constrained information transmission wide band, traditional uniform compression ratio (CR) algorithms are unable to consider both the requirements of high CR and image reconstruction quality. The non-uniform image compression methods based on regions of interest (ROI) apply lossless or high-fidelity compression to ROI while adopting a high-CR method to the background, which results in the effective compression of image redundancies, with important information being preserved as much as possible.

Discriminating the ROI from the background is important. The most direct method to achieve this is to allow the user to select the interesting regions[1]. However, such an image-processing system may become quite complicated. Unsupervised extraction of a specific ROI from an image is an important procedure for computer vision and image-processing algorithms.

Human beings have the remarkable ability to determine ROI in complex scenes quickly. This ability is called "selective visual attention mechanism" (SVAM). Therefore, introducing SVAM in unsupervised ROI extraction is important and necessary because it can reduce computational complexity[2], save computational resources, and improve the efficiency of image processing.

In this letter, a novel, non-uniform image compression approach based on an unsupervised ROI extraction algorithm is proposed. The algorithm can successfully distinguish ROI from the original image. When ROI have been extracted, the ROI and other regions are encoded with different CRs[3] using a popular image compression algorithm.

The proposed approach adopts the intersection of the saliency map produced by a modified Itti-Koch visual attention model and the segmentation result of the clustering analysis algorithm to determine ROI. A flowchart of the proposed model is shown in Fig. 1.

If peaks in the saliency map overlap with regions determined by the image segmentation, the ROI is extracted based on these regions. One of the most important advantages of the proposed model is that its execution is entirely unsupervised.

The original Itti-Koch model of visual attention uses the dyadic Gaussian pyramid to subsample the input image[4]. In contrast, in the proposed modified model, the wavelet[5] is adopted because of its multiresolution image representation, which is in accordance with human visual characteristics. Six different spatial scales are created. Each level is then decomposed into seven feature channels (one for intensity, two for color, and four for orientation).

If r, g, and b represent the red, green, and blue values of the color image, the intensity feature I and orientation feature $O(\theta)$ (with orientation 0°, 45°, 90°, and 135°, respectively) are computed as described in the Itti-Koch model[6]. The red-green (RG) and blue-yellow (BY) color opponencies are defined as[7]

$$\mathrm{RG} = \frac{\mathrm{r} - \mathrm{g}}{\max(\mathrm{r}, \mathrm{g}, \mathrm{b})}, \tag{1}$$

$$\mathrm{BY} = \frac{\mathrm{b} - \min(\mathrm{r}, \mathrm{g})}{\max(\mathrm{r}, \mathrm{g}, \mathrm{b})}. \tag{2}$$

To avoid large fluctuations of color opponency values at low luminance, RG and BY are set to zero at locations with $\max(\mathrm{r}, \mathrm{g}, \mathrm{b}) < 0.1$, assuming a dynamic range

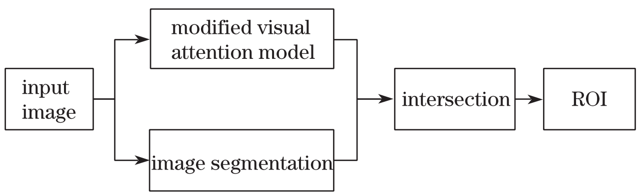


Fig. 1. Flowchart of ROI extraction.

of [0,1]. The definitions in Eqs. (1) and (2) are very deviant from the original model by Itti *et al*.

Center-surround receptive fields are simulated by across-scale subtraction ($\ominus$) between features at the center (c) and the surround (s) levels of the different spatial scales[6], yielding the following "feature maps":

$$\mathrm{FM}_{c,s} = |f(c) \ominus f(s)| \quad \forall f \in F = F_{\mathrm{I}} \cup F_{\mathrm{C}} \cup F_{\mathrm{O}}, \tag{3}$$

with

$$F_{\mathrm{I}} = \{I\}, \;\; F_{\mathrm{C}} = \{\mathrm{RG}, \mathrm{BY}\}, \;\; F_{\mathrm{O}} = O\{0^\circ, 45^\circ, 90^\circ, 135^\circ\}. \tag{4}$$

When feature maps are generated, the next step is to combine these with a saliency map based on the hypothesis that similar features compete for saliency, whereas different features independently contribute to the saliency map[6]. As a result, a normalization operator $N(\cdot)$ is needed for the combination of the saliency map.

The original Itti-Koch model normalizes each feature map to a fixed range and sums up all the maps, which results in poor performance in complex scenes[1]. In this letter, an iterative strategy based on the two-dimensional (2D) difference of Gaussians (DoG) filter is adopted to realize the local normalization of difference feature maps. The DoG filter yields strong local excitation at each visual location, which is counteracted by broad inhibition from neighboring locations[7].

Specifically, the definition of the DoG function is

$$\mathrm{DoG}(x,y) = \frac{c_{\mathrm{ex}}^2}{2\pi\sigma_{\mathrm{ex}}^2}\mathrm{e}^{-\frac{x^2+y^2}{2\sigma_{\mathrm{ex}}^2}} - \frac{c_{\mathrm{inh}}^2}{2\pi\sigma_{\mathrm{inh}}^2}\mathrm{e}^{-\frac{x^2+y^2}{2\sigma_{\mathrm{inh}}^2}}. \tag{5}$$

In the present implementation, σ_{ex} and σ_{inh} are respectively 2% and 25% of the input image width[9], and $c_{\mathrm{ex}} = 0.5$ and $c_{\mathrm{inh}} = 1.5$. At each iteration process, the feature map FM is subjected to the following operation:

$$N(\mathrm{FM}) \leftarrow |\mathrm{FM} + \mathrm{FM} * \mathrm{DoG} - C_{\mathrm{inh}}|_{\geqslant 0}, \tag{6}$$

whereas $C_{\mathrm{inh}} = 0.02$, with the feature maps initially scaled between 0 and 1. Additional data regarding the iteration procedure with DoG are detailedly described in Ref. [9].

The normalized feature maps are summed up across scales to generate the following conspicuity maps: I_{C} for intensity, C_{C} for color, and O_{C} for orientation.

$$I_{\mathrm{C}} = \bigoplus_{c=0}^{2} \bigoplus_{s=c+2}^{c+3} N[I(c,s)], \tag{7}$$

$$C_{\mathrm{C}} = \bigoplus_{c=0}^{2} \bigoplus_{s=c+2}^{c+3} \{N[\mathrm{RG}(c,s)] + N[\mathrm{BY}(c,s)]\}, \tag{8}$$

$$O_{\mathrm{C}} = \sum_{\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}} N\Big\{\bigoplus_{c=0}^{2} \bigoplus_{s=c+2}^{c+3} N[O(c,s,\theta)]\Big\}. \tag{9}$$

Finally, the conspicuity maps are normalized and summed into the saliency map as

$$S = \omega_{\mathrm{i}}[N(I_{\mathrm{C}})] + \omega_{\mathrm{c}}[N(C_{\mathrm{C}})] + \omega_{\mathrm{o}}[N(O_{\mathrm{C}})], \tag{10}$$

where ω_{i}, ω_{c}, and ω_{o} are weights for the combination of these conspicuity maps. These weights may be set to different values when top-down visual attention is involved and a specific task is given. In this letter, all three weight coefficients are set to 1/3.

Figure 2 shows the comparison of saliency map results generated by the Itti's model and the proposed approach. The saliency map results generated by the proposed approach are obviously sparser than those generated by the Itti's model, and are thus more easily adopted for the saliency decision.

Color is an important and useful cue in extracting information from an image. The input image is first converted from a RGB color space to a Lab color space. Unlike the RGB color models, Lab color is designed to approximate human vision, which means that distances in this space are consistent with human perception[10].

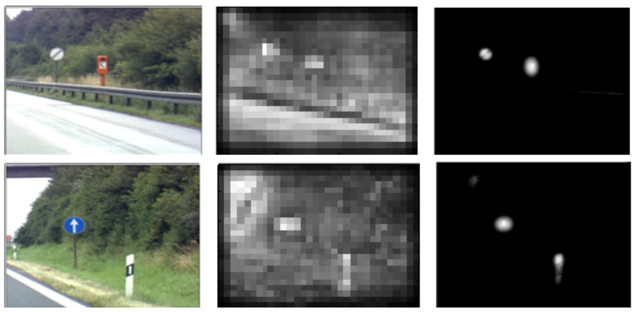
After the color space conversion, the k-means clustering analysis algorithm is adopted to conduct the color image segmentation[11]. Given a set of observations $(x_1, x_2, \cdots, x_n)$, where each observation is a d-dimensional real vector, this algorithm aims to partition the n observations into k ($k \leqslant n$) sets $O = \{O_1, O_2, \cdots, O_k\}$, so as to minimize the within-cluster sum of squares as

$$\arg\min_{O} = \sum_{i=1}^{k} \sum_{x_j \in O_i} ||x_j - c_i||^2. \tag{11}$$

In this expression, $||x_j - c_i||^2$ is a chosen distance measure between a data point x_j and the cluster center c_i, and $||x_j - c_i||^2$ is a distance indicator of the n data points from their respective cluster centers[12].

When all the pixels have been assigned to certain regions of the image, one of the cluster regions is reserved and the others are set to zero, as shown in Fig. 3. In the segmentation result, each color represents a coherent region of the original image.

The second stage of the proposed model generates ROI that correspond to the most salient areas of the image. This proposed model is inspired by the approach used by Rutishauser *et al.*[1]. The algorithm described in this letter for ROI extraction combines the saliency map produced by the modified Itti-Koch model with the segmentation result of the clustering analysis to leverage the strengths of either approach without suffering from their shortcomings[4]. The proposed model appreciates not only the magnitude of the peaks in the saliency map,



(a) (b) (c)

Fig. 2. Comparison of saliency map results generated by the Itti's model and the proposed approach. (a) Original image from Ref. [13]; (b) and (c) Saliency maps by Itti's model and our approach.

but also the size and contour of the regions generated by image segmentation.

Figure 4 shows the comparison of the ROI extraction results by Itti's model (labeled by the yellow circles) and the proposed method (all of these processes are implemented by Matlab using functions). The proposed approach results in precision and correctness of the unsupervised ROI extraction. The extracted ROI are considered to be the basis for the next step, non-uniform image compression.

The non-uniform image compression approach is based on the separation of the treatment of the ROI with that of the context. This means that images can be compressed with different accuracy in different regions, which can code the important parts inside the image with better quality than the rest of the contents. Figure 5 presents additional details of the non-uniform image compression approach based on ROI.

The JPEG image compression algorithm is selected to encode the ROI and context of the input image with separate quality factors (Q factors), which are numerical values that determine CRs in the compression process. The comparison of uniform and non-uniform image compression results are shown in Fig. 6. Figure 6(a)

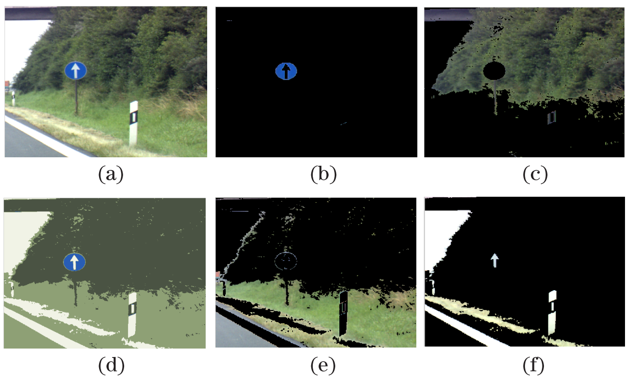


Fig. 3. Result of the regions segmented by the clustering algorithm. (a) Original image; (d) segmentation result; (b), (c), (e), and (f) objects in clusters 1, 2, 3, and 4.

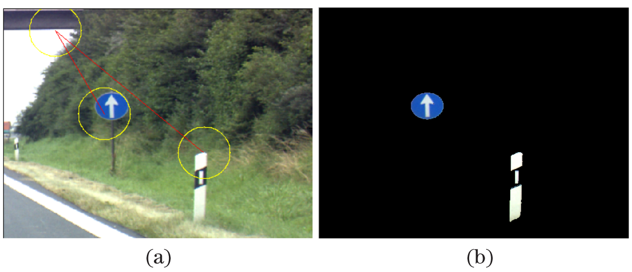


Fig. 4. Results of ROI extraction. (a) ROI labeled by Itti's model and (b) ROI extracted by our method.

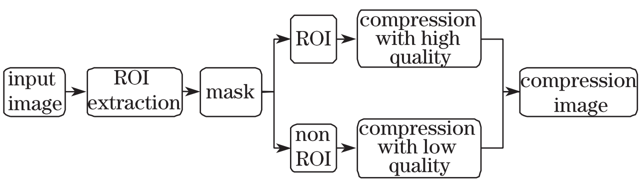


Fig. 5. Non-uniform image compression approach: detailed block diagram.

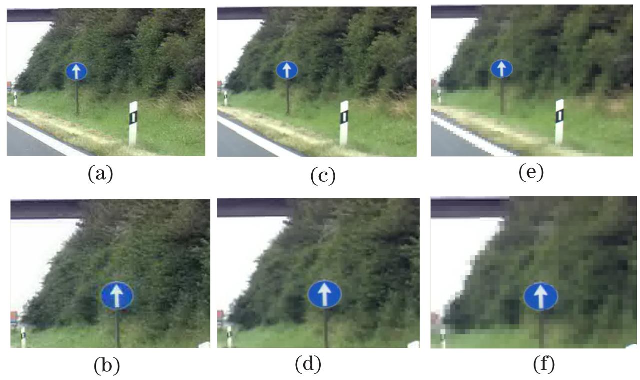


Fig. 6. Comparison of uniform and non-uniform image compression results. (a) Uniform JPEG compression (CR=66.25); (c) and (e) non-uniform compression based on ROI with CR=61.11 and 112.21; (b), (d), and (f) local amplification effect of (a), (c), and (e).

Table 1. Evaluation of the Compression Effect

Fig. 6	File Size	CR	PSNR	
			(N_ROI)	(ROI)
(a)	8 903 B	66.25	31.48	31.48
(c)	9 654 B	61.11	30.43	35.77
(e)	5 257 B	112.21	28.81	35.77

The file size of the original image (.bmp) is 589 878 B.

is the result of traditional uniform image compression and Figs. 6(c) and (e) are the results generated by the proposed non-uniform image compression algorithm. In Figs. 6 (c) and (e), the ROI extracted are treated with the same quality factor, whereas other regions are dealt with lower quality factors.

Figures 6(b), (d), and (f) are the local amplification effect diagrams. The non-uniform image compression approach proposed in this letter obviously has a better performance in ROI image quality compared with the uniform image compression, whereas separate CRs have distinct results.

The evaluation of the proposed compression approach is measured by CR and peak signal-to-noise ratio (PSNR), whereas the PSNR of the color image is regarded as the average PSNR of the r, g, and b components. Table 1 shows the comparison of the uniform JPEG compression method and the proposed non-uniform image compression method with distinct CRs.

As shown in the table, the proposed non-uniform image compression method achieves a better compression effect in ROI because of a slight decrease in the CR with the cost of image degradation in the non-ROI.

In conclusion, a non-uniform image compression method based on an unsupervised ROI extraction algorithm is proposed. The unsupervised ROI extraction algorithm adopts the intersection of the saliency map produced by the proposed modified classical Itti-Koch visual attention model and the succeeding clustering segmentation result, which provides a better performance in the ROI extraction precision and correctness.

The applications used in the proposed reformative visual attention model include the wavelet for the image down-sampling, new definitions of the color opponencies, and a local iterative method based on 2D DoG filter for the feature combination strategy. Adaptive non-uniform image compression method based on the proposed ROI extraction method results in a better compression effect in ROI because of a slight decrease in the CR compared with other traditional approaches. Future work includes the refinement on the extraction of ROI and the extension of the proposed visual attention model to incorporate a top-down component. In addition, a new ROI coding strategy should also be taken into account in future research because the DCT adopted in the JPEG image compression algorithm may result in a block effect at low bit rates.

This work was supported by the National Natural Science Foundation of China (No. 60702017) and the Scientific Research Fund of Key Laboratory of Photoelectronic Imaging Technology and System, Ministry of Education of China (No. 2010OEIOF02).

References

1. U. Rutishauser, D. Walther, C. Koch, and P. Perona, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (2004).
2. Q. Zhang, G. Gu, and H. Xiao, J. Multimed. **4,** 363 (2009).
3. L. Zhang and K. Wang, Chin. Opt. Lett. **4,** 76 (2006).
4. L. Itti, and C. Koch, Nat. Rev. Neurosci. **2,** 194 (2001).
5. S. G. Mallat, IEEE Trans. Pattern Anal. Mach. Intell. **11,** 674 (1989).
6. L. Itti, C. Koch, and E. Niebur, IEEE Trans. Pattern Anal. Mach. Intell. **20,** 1254 (1998).
7. D. Walther, "Interactions of visual attention and object recognition: computational modeling, algorithms, and psychophysics" PhD. Thesis (California Institute of Technology, 2006).
8. J. G. Daugman, J. Opt. Soc. A **2,** 1160 (1985).
9. L. Itti and C. Koch, Vis. Res. **40,** 1489 (2000).
10. Y. Lu, X. Zhang, J. Kong, X. Wang, and J. Zhang, in *Proceedings of Congress on Image and Signal Processing* 339 (2008).
11. J. Lan and K. Shen, Chin. Opt. Lett. **8,** 286 (2010).
12. X. Bian, T. Zhang, and X. Zhang, Chin. Opt. Lett. **9,** 011002 (2011).
13. L. Itti and C. Koch, "iLab Image Databases", http://ilab.usc.edu/imgdbs/ (2002).