Adaptive Color Image Compression based on Visual Attention

Nabil Ouerhani, Javier Bracamonte, Heinz Hügli, Michael Ansorge, Fausto Pellandini
Institute of Microtechnology, University of Neuchâtel
Rue A.-L. Breguet 2, CH-2000 Neuchâtel, Switzerland
{*Nabil.Ouerhani, Javier.Bracamonte*}*@unine.ch*

Abstract

This paper reports an adaptive still color image compression method which produces automatically-selected ROIs with a higher reconstruction quality with respect to the rest of the input image. The ROIs are generated on-the-fly with a purely data-driven technique based on visual attention. Inspired from biological vision, the multicue visual attention algorithm detects the most visually-salient regions of an image. Thus, when operating in systems with low bit-rate constraints, the adaptive coding scheme favors the allocation of a higher number of bits to those image regions that are more conspicuous to the human visual system. The compressed image files produced by this adaptive method are fully compatible with the JPEG standard, which favors their widespread utilization.

1 Introduction

Visual attention is the ability to rapidly detect interesting parts of a given scene. Using visual attention in a computer vision system permits a rapid selection of a subset of the available sensory information before further processing. The selected locations usually correlates with the conspicuous parts of the scene.

Various computational models of visual attention have been presented in previous works [1, 2]. These models are, in general, data-driven and based on the feature integration principle [9] which is inspired by psychophysical studies on human visual attention. Known as saliency-based, the model presented in [2] considers a variety of scene features (intensity, orientation, and color) to compute a set of conspicuity maps which are then combined into the final saliency map. The conspicuity operator is a kind of "contrast detector" which, applied on a feature map, brings out the regions of the scene containing salient information and thus, provides a set of relevant regions-of-interest, ROIs.

The availability of a set of preidentified ROIs can be associated with spatially adaptive image coding in order to obtain higher compression ratios while keeping a good reconstruction quality within the visually important regions of the image.

Previous works have dealt with the problem of the identification of ROIs to spatially adapt the compression according to the relative importance of regions [3, 7, 10]. A recent work [5, 6] presented an algorithm based on automatically preidentified ROIs which have been computed by means of a biologically plausible technique. This technique deals, however, only with grey scale images; chromatic features were not considered.

In this work, we investigate a different biologically inspired technique to identify visually relevant regions on an image. In contrast with the method discussed above, the multicue saliency-based model of visual attention combines chromatic as well as monochromatic features to identify ROIs on color images.

This paper is organized as follows. Section 2 presents the saliency-based model of visual attention. The adaptive image compression algorithm is described in Section 3. Section 4 reports the results of experiments involving a variety of color images, in order to assess the effectiveness of visual attention in the field of adaptive color image compression. Finally, the conclusions are stated in Section 5.

2 Visual attention model

2.1 Saliency-based model

According to a generally admitted model of visual perception [4], a visual attention task can be achieved in three main steps (Fig. 1).

1) First, a number (n) of features are extracted from the scene by computing the so-called feature maps. A feature map represents the image of the scene, based on a well-defined feature. This leads to a multi-feature representation of the scene. Five feature maps have been considered in our implementation: a) The difference between the red and the green component ($R - G$), b) The difference between the blue and the yellow component ($B - Y$), c) The intensity
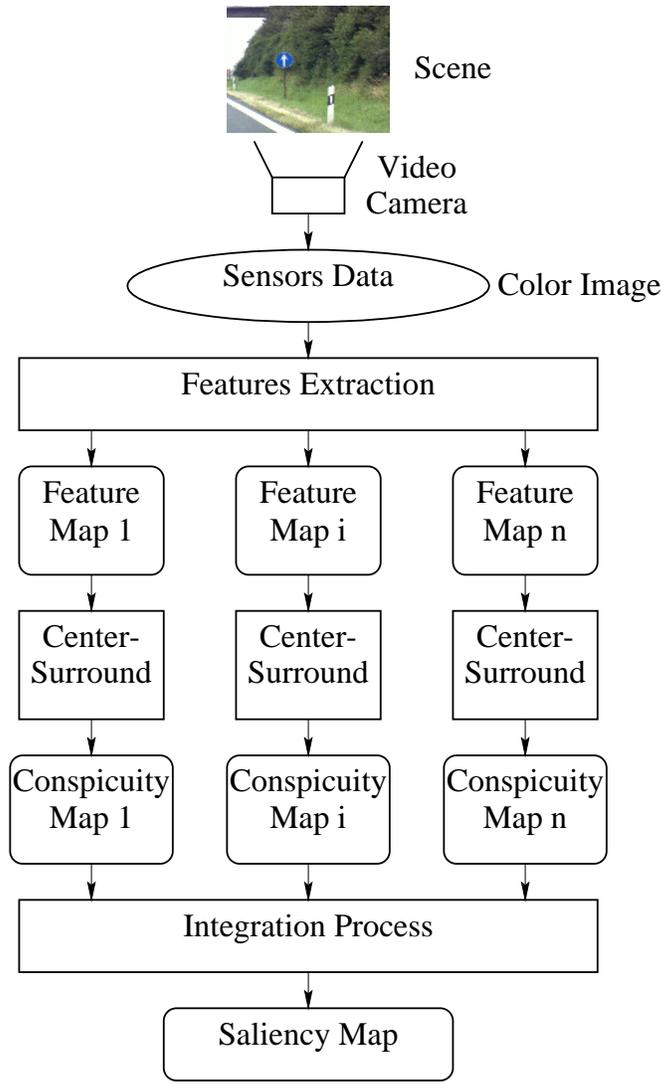
**Figure 1.** Scheme of a computational model of attention.

image, and d) The gradient orientation map.

2) In a second step, each feature map is transformed in its conspicuity map. Each conspicuity map highlights the parts of the scene that, according to a specific feature, strongly differ from its surrounding. In biologically plausible models, this is usually achieved by using a *center-surround-*mechanism. Practically, this mechanism can be implemented with a *difference-of-Gaussians*-filter, which can be applied on feature maps to extract local activities for each feature type.

3) In the last stage of the attention model, the n conspicuity maps C_i are integrated together, in a competitive way, into a *saliency map* $\mathcal{S}$ in accordance with equation 1.

$$\mathcal{S} = \sum_{i=1}^{n} w_i \mathcal{C}_i \qquad (1)$$

The competition between conspicuity maps is usually established by selecting weights w_i according to a weighting function w, like the one presented in [2]: $w = (M - \overline{m})^2$, where M is the maximum activity of the conspicuity map and $\overline{m}$ is the average of all its local maxima. w measures how the most active locations differ from the average. Thus, this weighting function promotes conspicuity maps in which a small number of strong peaks of activity is present. Maps that contain numerous comparable peak responses are demoted. It is obvious that this competitive mechanism is purely data-driven and does not require any a priori knowledge about the analyzed scene.
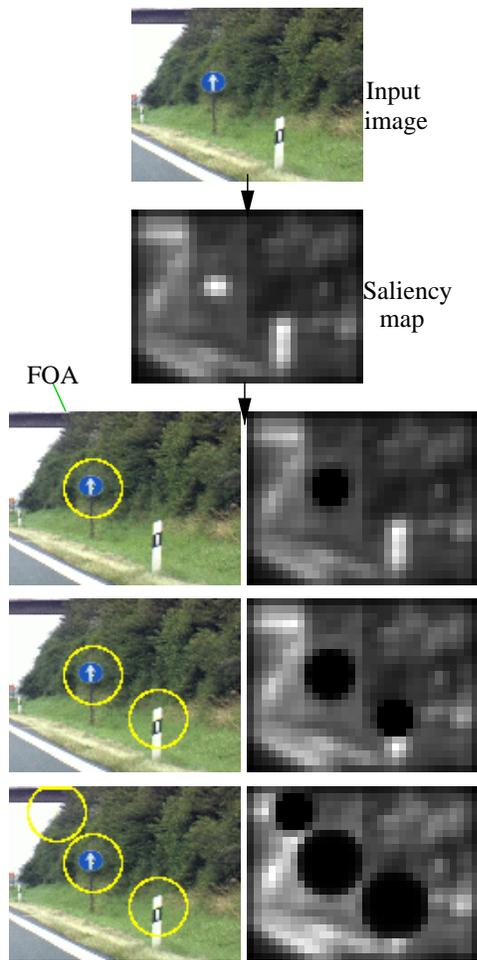
Figure 2. Salient regions selection. Applying a WTA mechanism to a saliency map permits the selection of the most conspicuous parts of the image.

2.2 Selection of salient locations

At any given time, the maximum of the saliency map defines the most salient location, to which the focus of attention (FOA) should be directed. A "winner-take-all" (WTA) mechanism [2] is used to detect, successively, the significant regions. Given a saliency map computed by the saliency-based model of visual attention, the WTA mechanism starts with selecting the location with the maximum value on the map. This selected region is considered as the most salient part of the image (winner). The FOA is then shifted to this location. Local inhibition is activated in the saliency map, in an area around the actual FOA. This yields dynamical shifts of the FOA by allowing the next most salient location to subsequently become the winner. Besides, the inhibition mechanism prevents the FOA from returning to previously attended locations. An example of
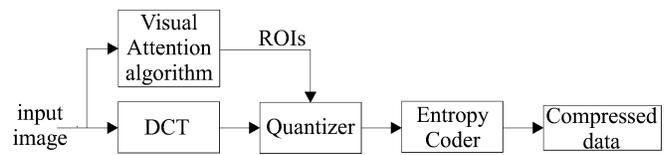
**Figure 3.** Adaptive JPEG-based Image compression Algorithm

salient regions selection based on the WTA mechanism is given on the Figure 2.

3 Adaptive compression algorithm

Figure 3 shows a block diagram of the adaptive coding method. This scheme follows the same operations of the baseline JPEG algorithm, albeit with a quantization unit that has been modified to receive an additional input: a binary image produced by the visual attention stage. This binary information will indicate the quantizer to execute either a short- or a large-step quantization of the DCT coefficients, depending on whether a given 8x8-pixel block lies inside or outside any of the identified ROIs.

The modified quantization unit uses the same normalization array $\mathbf{N}$ proposed in the official JPEG document [8]. This unit requires also two scale factor parameters: sf_0 and sf_1. These values can be permanently set by the user, or left to be set by the system in correspondence with a required compression bit rate for a given image. For those 8x8-pixel blocks with a majority of pixels lying within the ROIs, the quantization is executed using the normalization array previously scaled by sf_0. For the rest of the blocks, $\mathbf{N}$ is scaled by sf_1 before quantization. To preserve image detail within the ROIs, sf_0 is usually chosen to be in the interval [0.5, 1], while sf_1 is generally selected to be a real number larger than two.

The binary image that is produced by the visual attention stage and used by the quantization unit, represents overhead information to be embedded in the compressed data bitstream. This data is required for a decoder to execute the corresponding ROI-dependent inverse quantization of the DCT coefficients. The presence of this overhead data precludes the compressed image from being reconstructed
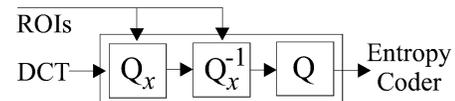
Figure 4. Quantizer of the adaptive compression algorithm to produce JPEG compatibility

(a) Original image

(b) Identified ROIs

(c) Standard JPEG

(d) Adaptive compression

Figure 5. Adaptive versus non-adaptive compression: Example 1.

by a standard JPEG decoder. If JPEG compatibility is not an issue, then any given JPEG decoder can be straightforwardly modified to accept the additional binary information, and accordingly execute the decoding operation of the compressed bitstream.

JPEG compatibility, however, could be required in a large number of systems, and it can be easily achieved in exchange of two additional quantization operations, as shown in Figure 4. After the DCT operation, the initial ROI-dependent quantization (Q_x) is followed by a corresponding ROI-dependent inverse quantization (Q_x^{-1}). After this point, the overhead data is no longer required and the current DCT coefficients can be re-normalized using a regular, ROI-independent, JPEG quantization (Q). This procedure produces a spatially adaptive compressed image which

is fully compatible with the JPEG standard. This was the scheme used to produce the results presented in Section 4.

4 Experiments

This section is dedicated to report experiments involving two sets of color images, in order to assess the usefulness of visual attention in the field of color image compression by means of the adaptive coding scheme presented in Section 3. For each example, a color image is acquired; using this image as input, the visual attention algorithm computes a set of ROIs. In these experiments the number of identified ROIs has been, for simplicity, limited to three. A yellow circle is drawn around each of the identified ROIs. Afterwards,

(a) Original image

(b) Identified ROIs

(c) Standard JPEG

(d) Adaptive compression

Figure 6. Adaptive versus non-adaptive compression: Example 2.

the color image is compressed using two methods, a) standard JPEG, and b) the JPEG-based adaptive compression algorithm. With both methods, and for all the experiments, the overall compression ratio produced was 100:1.

The images of the reported experiments are shown in Fig. 5 and 6, they mainly feature two persons facing the camera. Based on the considered features described in Section 2, the two persons stand out from the rest of the scene, and are thus, natural candidates for the ROIs, to be automatically identified by the visual attention algorithm. Figures 5b) and 6b) show, that as expected, the two persons' faces figure among the three most salient regions of the image. The adaptive compression algorithm takes into account the relative importance of these image regions. Consequently, the reconstructed images (bottom-right images in Fig. 5 and

6) preserve the visual details of the two faces, which may be relevant for the recognition of the two persons. On the other hand, the persons' faces have lost important perceptual details when using the standard JPEG method (bottom-left images in Fig. 5 and 6). In the latter case, one may have difficulty to identify the two persons.

The advantage of the adaptive algorithm is highlighted in Fig. 7, where the rightmost ROI has been zoomed in.

These examples clearly validate the adaptive color image compression algorithm based on visual attention. Despite the unavailability of any a priori knowledge about the analyzed images, the reported coding scheme permits the preservation of perceptually important image details.

(a) Standard JPEG   (b) Adaptive compression

Figure 7. Adaptive versus non-adaptive compression: Zoomed ROI from Example 2

5   Conclusion

After introducing the biologically inspired saliency-based model of visual attention which permits the identification of perceptually salient regions-of-interest on color images, this paper reported an adaptive still color image compression method. Based on automatically generated ROIs, this method favors the preservation of perceptually important image details. The compressed image files produced by this adaptive method are fully compatible with the JPEG standard, which favors their widespread utilization. The presented results clearly validate the reported compression method by showing its advantages over the non-adaptive JPEG algorithm.

The JPEG algorithm was chosen due to its fair computational complexity, but it is obvious that other methods (e.g., JPEG2000) could be equally used to compress the input images, in an adaptive manner, in accordance with the ROIs provided by the visual attention stage. Furthermore, the reported visual attention algorithm can be extended to detect ROIs in temporally changing scenes, by introducing motion as an additional scene feature into the model. This extension may permit spatially and temporally adaptive compression of video sequences.

Acknowledgments

References

[1] S. Culhane and J. Tsotsos. A prototype for data-driven visual attention. ICPR92, Vol. 1, pp. 36-40, Hague, The Netherlands, 1992.

[2] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 20(11), pp. 1254-1259, 1998.

[3] A. Kundu. Enhancement of JPEG coded images by adaptive spatial filtering. International Conference on Image processing, Vol. 1, pp. 23-26, 1995.

[4] R. Milanese. Detecting salient regions in an image: from biological evidence to computer implementation. Ph.D. Thesis, Dept. of Computer Science, University of Geneva, Switzerland, Dec. 1993.

[5] C. Privitera and L. Stark. Focused JPEG encoding based upon automatic preidentified regions of interests. Proc. SPIE, Vol. 3 644, pp. 552-558, 1999.

[6] C. Privitera and L. Stark. Algorithms for defining visual regions-of-interest: Comparison with eye fixations. Pattern Analysis and Mashine Intelligence (PAMI), Vol. 22, Nr. 9, pp. 970-981, September 2000.

[7] R. Rosenholtz and A. Watson. Perceptual adaptive JPEG coding. Proceedings, IEEE International Conference on Image Processing, Lausanne, Switzerland, Vol. 1, pp. 901-904, 1996.

[8] ITU-T Recommendation T.81. Digital compression and coding of continuous-tone still images. September, 1992.

[9] A. Treisman and G. Gelade. A feature-integration theory of attention. Cognitive Psychology, pp. 97-136, 12, 1980.

[10] J. Zhao, Y. Shimazu, K. Ohta, R. Hayasaka, and Y. Matsuschita. A JPEG codec adaptive to the relative importance of regions in an image. Transaction of information Processing Society of Japan vol. 38(8), pp. 1531-1542, 1997.