SPARE TIME ACTIVITY SHEETS FROM PHOTO ALBUMS

Gabriela Csurka and Marco Bressan

Xerox Research Centre Europe, 6, ch. de Maupertuis, 38240 Meylan, France

Keywords: Non-photorealistic rendering, Semantic segmentation, Drawings, Coloring and activity page.

Abstract: We present a technique to generate some popular activity sheets from arbitrary images, in particular user photographs. We focus on activity sheets that are closely linked to coloring and shape completion. We first introduce a baseline approach based on color regions that works well for cartoon-like images and uncluttered photographs. In more complex scenes, we show how this approach can be integrated with global textural cues for increasing the level of details that can convey semantic information. A final local stage takes advantage of object recognition and scene classification techniques for selective detailing in the foreground background regions. Though the resulting approach can be deployed in a fully automatic fashion, interactivity can be a desirable feature since it allows to account for errors and, more important, increase the level of personalization. We propose three levels of interactivity, depending on the user skills. For all steps of our system and addressed activity sheets we show representative results.

1 INTRODUCTION

Children enjoy coloring. In addition, children like browsing their own family albums and looking at their own photos. If given the option, children will prefer to select the pictures they want to color, e.g. characters from their favorite cartoons, images from a particular subject they find on the internet, personal family photos, etc.

Coloring images are generally simple black and white silhouette or border images with well separated regions, each corresponding to a different color. These images can also present several differences in style (see Figure 1) leading to different spare time activities such as unsupervised silhouette coloring, numbered/labeled region coloring, dot linking, etc. They are also often used in kindergarten and elementary schools, where the labeling has to be deduced as part of an exercise e.g. a mathematical formula.

Typically, most of these activities were available on cartoon-like images and drawings. Transforming a printed photograph into a drawing suitable for any of these activities requires a complex manual process with multiple steps. With the popularity of digital photography, it is natural to device a digital technique to simplify this process.

The main challenge we address in this paper is to obtain coloring pages from the arbitrary types of images children might be interested in coloring or filling

i.e. photographs and cartoons. We propose a set of automatic and semi-automatic tools based on state-of-the-art image analysis and processing techniques that allow a non-expert user to generate quickly these spare-time activity sheets from arbitrary images, particularly user photographs. These tools can be easily plugged-in any interactive photo-editing system, can be added to online coloring services or can be part of a photographic print flow.

The coloring page creation can be seen as a particular case and application of photographic stylization and abstraction. Indeed, they share many common components, such as building edge maps and image segmentation, even face detection for non-realistic rendering (Brooks, 2007), and therefore these components used or developed in the former field can be re-used to inspire and to improve the coloring page creation. The former techniques are not specifically adapted to the coloring page creation and even less to derive the divers activity sheets. Their aim is to obtain painting rendering of the images (DeCarlo and Santella, 2002; du Buf et al., 2006; Olmos and Kingdom, 2006), stained glass effect (Mould, 2003; Brooks, 2006) or to enhance the compression rate for visual communication efficiency (Winnemöller et al., 2006). They generally combine the luminance edge maps with the "abstracted colors" of the image which mutually compensate the visual imperfection of both of them, leading to a paintings rendering stylized ef-
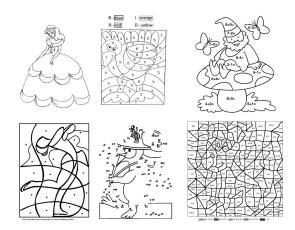
Figure 1: Different types of activity sheets.

fect.

We propose a method to generate coloring book pages automatically from user photos. The only other approach we are aware of with this objective is the commercial product Kidware Photo Color Tool (http://www.kidware.net). The results of the latter system are simple weighted edge maps similar to our texture edges (see section 3.1 and Figure 3), its main drawback beeing that many of the regions are not closed. As coloring images are mostly designed for young children, simplicity is desirable. Preferred images are black and white silhouette or border images with well separated regions, each corresponding to a different color. Another advantage of our system compared to it is that we address a wider range of derived activities (see section 4).

The paper is organized as follows. In section 2, we describe a simple and basic method which works well for simple images containing well contrasted and uncluttered objects with relatively uniform background. In section 3 we extend this method in order to cope with more complex scenes. In section 4 we describe methods to obtain diverse activity sheets. Section 5 discusses multiple levels of interactivity levels and Section 6 concludes the paper.

2 THE BASIC SYSTEM

We first propose a relatively simple system to obtain automatically coloring pages from a natural or cartoon-like image which nevertheless goes beyond a simple weighted edge map. This approach is suitable for images where regions are easily distinguishable and each region shows fairly uniform colors. Figure 2 shows a few examples obtained on real images with this basic system.

The main steps of this system are:

Figure 2: Example results for the basic approach.

1. *Color Conversion*

   We first transform the image from RGB space to some chrominance-luminance space. We choose L^*ab, as the Euclidean distance in this space has perceptual interpretation which can be of advantage for metric-based processing such as clustering.

2. *Edge-Preserving Low-pass Filtering*

   Next, we apply an edge-preserving low-pass filter (EPLP) to the different channels of the image. This seeks to reduce image noise which can lead to extra edges or image segments non-relevant for further processing. Simple median filtering can be used, or some more sophisticated methods such as edge-preserving maximum homogeneity neighbour filtering (Garnica et al., 2000) or anisotropic diffusion filtering (Perona and Malik, 1991).

3. *Image Segmentation*

   The third step consists in low level image segmentation or region clustering. The most common approaches are based on Normalized Cut (Jianbo and Jitendra, 2000) or Mean Shift (Comaniciu and Meer, 2002). These two methods in particular take into account the spatial closeness of the pixels and therefore lead to more compact segments. In our experiments, we used Mean Shift with a flat kernel and low color and spatial bandwidths ($\sigma_s, \sigma_r \in [5, 10]$)). The bandwidth parameter allows handling the coarseness similarly in different images without specifying the exact number of clusters to be found in the image. In order to ensure we do not miss any perceptually important boundary, we intentionally use a low bandwidth to over-segment the image.

4. *Region Merging*

   As we intentionally over-segment the image, in the fourth step, we do a region merging based on a set of rules that take into account both spatial and perceptual information. The merging criterion is different from the measure used by the meanshift,

meaning that the region merging leads to a different results than using higher bandwidth would do. The rules are very simple:

(a) If the area of the region is below a given threshold ($T_1 =0.0005\%$ of the image area), it will be absorbed by the most similar neighbor, independently of the color difference between them.

(b) If the area of the region is above T_1 but below a second threshold $T_2 > T_1$ ($T_2 =0.05\%$ of the image area) the region is merged with its most similar neighbor only if their color similarity is below a threshold that depends on the color variance of the image.

(c) If the area of the region is above T_2 the region is kept unchanged.

In both cases the color similarity is computed as a combination of distances in chrominance and luminance space, giving a higher weight (importance) to the distance in the chrominance space. This merging algorithm is applied iteratively until no modification is made or the maximum number of iterations is achieved. Alternatively, more complex region merging criteria could be applied, such as minimal cost edge removal in the corresponding region adjacency graph (Haris et al., 1998).

5. *Edge Detection*

Finally, we extract the closed edges of the obtained regions and use a morphological dilation function to get thicker region borders.

3 VISUAL ENHANCEMENT

The proposed basic system is a simple approach that gives satisfactory results in many cases. However, as two of the images in the second column of Figure 3 show, it can lead to less satisfactory results as scenes become more complex. In this section, we propose a few extensions to the basic system to improve the quality of the coloring pages.

3.1 Adding Texture Edges

One of the main difficulties of obtaining an acceptable coloring page for a complex scene images is that generally there are several objects/elements of the scene for which the level of "interesting" details can vary a lot. In a coloring page application, not all details will require the same level of attentio, e.g. a human face or the leaves or branches of a tree. An automatic system that has no knowledge about the image content will

Figure 3: Example results for the system with addition of texture edges. In second column the results of the basic system B, in the third column the edge map obtained by the DoG algorithm normalized to [0,1] and in the last column the weighted combination of them.

handle these regions in a similar fashion. The global parameters can be tuned to increase or decrease details, but the same criterion will be applied to all regions.

In order to handle this, we propose a first solution based on texture/luminance edges. To extract the luminance edges we use the Difference of Gaussians (DoG) algorithm as it approximates well the Laplacian of Gaussian, known to lead to a weighted edge map. Furthermore, the DoG is believed to mimic well how neural processing in the retina of the eye extracts details from images destined for transmission to the brain. After elimination of small edges and isolated dots in the DoG map, these luminance edges are combined with the region boundaries obtained in section 2. We used a weighted combination giving a higher weight[1] to the original coloring page borders in order to let them guide the coloring.

The main role of this combination is visual enhancement, however in many cases it can also compensate for missing borders between regions that were wrongly merged either by the low level segmentation or by the region merging step. These cues further help children to better understand the content if they do not have the original model. Figure 3 shows a few examples of coloring pages with and without adding these texture edges. The final results are still not perfect compared to what a human would do manually, however it seems that children can cope well with these small imperfections (see Figure 10).

---

[1]In our experiments we used $max(E,DoG^\lambda)$ with $\lambda = 0.4$, where the original edge map E is binary and DoG was normalized to have values between [0,1].

Figure 4: Example results for the system with addition of skin detection and filling with original content.
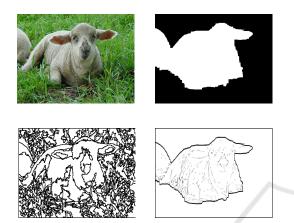

Figure 5: The sheep class mask (2nd image) obtained by CBIS was used to eliminate the background from the coloring page obtained by the basic system (3rd image). Furthermore, the low level segmentation was replaced by the high level masks boundary and texture edges added inside the relevant region (4th image).

Extracting ridges and valleys (Tran and Lux, 2004) can provide an alternative or a complement to this approach as artists frequently take advantage of ridges and valleys in their work.

3.2 Semantic Content Analyses

It is clear that if the system has some further knowledge about the semantic content of the regions then it can automatically handle those regions accordingly. For example, it can increase the weights or the thickness of the object's border, merge regions within the same semantic region, add luminance edges or ridges only to the regions of interest, etc. We can also alternatively fill some of the regions with the original content as in Figure 4 based on human skin detection.

In the last few years there were many publications and an increasing interest on semantic segmentation of images, i.e. assigning each pixel in an image to one of a set of predefined semantic classes. This is a supervised learning problem in contrast to the "classical" unsupervised low-level segmentation. There are three main groups of techniques that can be integrated with the system we propose:

1. *Foreground-background Separation*

   The objective here is to separate some foreground object or region of interest (ROI) from the background, not necessarily knowing what the object is. In most approaches proposed in the literature, the algorithm requires manual initialization, which can be simple enough to be done by a child (e.g. drawing a box or a contour around the ROI). In a pure automatic case, the system can either assume that the ROI is in the center of the image or use Visual Attention Maps (Itti et al., 1998) to initialize. After initialization, to get the foreground/background separation one might use image matting techniques (Sun et al., 2004), active contours (Juan et al., 2006) or to apply GrabCut (Rother et al., 2004).

2. *Object Detection and Localization*

   The main idea is to simultaneously recognize and segment out a predefined object class such as person, car, horse, etc. The techniques that address this problem (Winn and Jojic, 2005; Levin and Weiss, 2006) require that the set of object classes be predefined. In general, the system has to be trained with fairly clean examplary images. Faces and human skin are of particular interest in coloring page because users are more sensible concerning the results of segmentation or edges on a face or human body than on any other object. Much of the state of the art focuses on these categories: see (Viola and Jones, 2001; Yang et al., 2002) for face and (Vezhnevets et al., 2003; Tomaz et al., 2003) for human skin detection. As above, the initial detection results can again be further refined by matting or active contours to get a better object/background segmentation.

3. *Semantic based image Segmentation*

   In contrast to the two previous cases, in this approach the image can be partitioned generally in more than two semantically labeled (meaningful) regions. Of course the previous cases can be seen as particular instances of the semantic segmentation problem where only two classes are defined. Several techniques were recently proposed to solve this problem (Shotton et al., 2006; Yang et al., 2007; Csurka and Perronnin, 2008). We used the last one in our experiments (called CBIS in which follows).

The integration of these techniques with the coloring page system can hence further enhance the visual quality of the coloring page (see Figures 4 and 5), but they are of particular interest for the derived activity sheets as we will see.
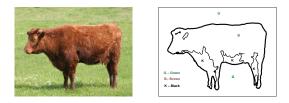
Figure 6: An example of a labeled coloring page.

4 DIVERSE ACTIVITY SHEETS

4.1 Region Labeling

In contrast to the unsupervised case, the idea here is that the child has to follow some rules to color each region. It can be simply the recognition of some letters as in the 2nd example of Figure 1), or it can be more complex such as mathematical or logical formulas. The latter are often used in kindergarten and elementary school with pedagogical purposes.

With our system, this can be done automatically, because (1) we have closed regions and (2) we have a representative color of each region (mean color, cluster center or mean shift mode). We can therefore select a set of standard colors (e.g. using the well-known, standard NBS-ISCC color name dictionary http://www.anthus.com/Colors/NBS.html), to find for each region the selected standard color which is closest to its representative color and plot the corresponding letters, formulas, shape, etc (depending on the children's age) onto it. Eventually, on the border or next to the image, the legend is printed with the labels and the corresponding color (see Figure 6).

4.2 Link the Dots

A second popular activity sheet example is the link-the-dots sheet (see 5th example in Figure 1). These sheets are also often used by kindergarten as they help children to learn number ordering and the alphabet.

The main idea is to take a single object boundary using one of the techniques described in section 3.2, sample dots on it, label them with letters, numbers or formulas following the contour and eventually delete the original contour.

The dots on the boundary can be sampled uniformly or obtained by more complex algorithms that seek for corners and inflexions points such as chain code detection (Liu and Srinath, 1990), local contours (Reche et al., 2002), direct estimation of the curve and its high curvature points (Chetverikov and Szabo, 1999; Hermann and Klette, 2005). We used a
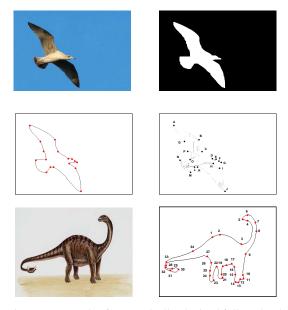

Figure 7: Example of automatically obtained follow-the-dot examples. In the first case we used the bird class mask obtained by CBIS and the dots were obtained by CSS corner detector on the boundary of the mask. We also added the DoG edges inside the object region to enhance the final result. In the second case, we initialized the GrabCut (Rother et al., 2004) with a box centered in the middle of the image and made foreground/background separation. We show the original object boundary on the results for pure visualization purposes.

local implementation of the popular method, the Curvature Scale Space (CSS) based corner detector (Abbasi et al., 1999; He and Yung, 2004) to obtain a set of dots on the object boundary (see Figure 7).

4.3 Object Discovery through Coloring

Finally, a third type activity sheet is to discover hidden objects through coloring. The objective can be a simple foreground or background coloring, labeled by dots as in 4th example in Figure 1; or more complex where a set of colors has to be used and the individual labels are mathematical or logical formulas (as in 6th example of Figure 1).

These sheets can also be derived from our solution when we have the knowledge of foreground/background or alternatively the semantic regions (see section 3.2). In these cases, the idea is to either use the over-segmentation we already have in step 3 (section 2) or we can combine the high level segmentations with some random partitioning of the image. Finally, the dots/formulas can be added automatically to sub-regions according to their semantic meanings[2]. We could also use forground/background

---

[2]In our experiments (see Figure 8) we obtained the se-

Figure 9: Example results for the system interactive region filling with original content.
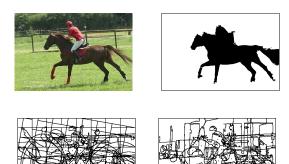
Figure 8: Example results for hidden object sheet. We used two different image partitionning strategy (many others can be used): partitionning the image by random parralel lines and random ellipses (third image) and respectively combining a few existent coloring pages as we have closed regions for them (fourth image). In both cases the dots were added to the regions that had a minimum of 70% overlap with the estimated "horse + person" mask (second image).

detection as described in section 3.2.

5 INTERACTIVITY

Application of the described steps, using the default parameters and having pre-selected the output style, results in a fully automatic coloring image generator. Alternatively, we can envisage the integration of the system with any photo editing tool or interactive coloring systems by adding different interactivity levels to the system:

1. At a first level of interaction the system allows the user (child) to select or to upload a photo. The photo is automatically processed and a set of coloring pages (with or without texture edges) and activity sheets are proposed.

2. A second level of interaction would be designed for older children or parents allowing to modify/adjust some of the parameters of the system. However, the interaction with the parameters has to be user friendly, such as choosing between less or more details, thiner or thicker edges, adding or not texture edges, adding letters or formulas, etc. Then the corresponding interior parameters are adjusted accordingly.
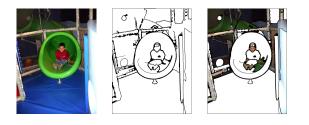
3. Finally a highest level of interaction could allow

matic meanings of the regions with the CBIS, considering all Pascal VOC 2007 classes as relevant regions (see http://pascallin.ecs.soton.ac.uk/challenges/VOC for further details).

the user to edit the obtained coloring page by some interactive tools such as:

- *Erasing Tool:* to delete selected edges (the two regions separated by the selected edges will be merged automatically).

- *Dot Adding Tool:* to complement follow-the-dot pages

- *Region Filling Tool:* to fill the region either by the texture edges, the mean color value or original content[3] (see Figure 9).

6 CONCLUSIONS

We propose a system that partially or totally automates the creation of a range of spare time activity sheets from photo albums. We provide solutions for unsupervised coloring page creation but also show how we can derive related activity sheets. The main originalities and advantages of the proposed system are that the original image is arbitrary, and that the coloring pages and derived activity sheets can be created automatically. We also propose different levels of interactivity that depend on the skill requirements we wish to impose.

The approach is simple and hence can be easily integrated in any photo editing or online coloring system. As there is no real groundtruth neither benchmark data it is difficult to establish the best parameter set of the system. In our experiments we processed hundreds of images of the Pascal VOC 2007 Challenge as we had for them the CBIS estimates and visually compared them to tune the examplary parameters reported and used to obtain the images shown in the paper. They are probably not the best choises, and,

---

[3]The main goal of such interaction can be visual quality enhancement but also can provide extra fun to the children. Indeed, in the case of interactive coloring of the page this can also be seen as a "*magical pencil*" that allows the child to fill image regions with the original content of the image instead of coloring it.

as a future work, we intend to do intensive preference user studies to better establish these parameters and also to compare different alternatives of the system.

We provided gave a set of automatically created color pages to a few children (see some of them in Figure 10) who accepted[4] and enjoyed coloring them for us.

REFERENCES

Abbasi, S., Mokhtarian, F., and Kittler, J. (1999). Curvature scale space image in shape similarity retrieval. *Multimedia Systems*, 7.

Brooks, S. (2006). Image-based stained glass. *IEEE Transactions on Visualization and Computer Graphics*, 6(12).

Brooks, S. (2007). Mixed media painting and portraiture. *IEEE Transactions on Visualization and Computer Graphics*, 5(13).

Chetverikov, D. and Szabo, Z. (1999). A simple and efficient algorithm for detection of high curvature points in planar curves. In *Workshop of Austrian Pattern Recognition Group*.

Comaniciu, D. and Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *PAMI*, 24.

Csurka, G. and Perronnin, F. (2008). Object class localization and semantic class based image segmentation. In *BMVC*.

DeCarlo, D. and Santella, A. (2002). Stylization and abstraction of photographs. In *SIGGRAPH*.

du Buf, H., Rodrigues, J., Nunes, S., Almeida, D., Brito, V., and Carvalho, J. (2006). Painterly rendering using human vision. In *VIRTUAL, Advances in Computer Graphics*.

Garnica, C., Boochs, F., and Twardochlib, M. (2000). A new approach to edge-preserving smoothing for edge extraction and image segmentation. In *ISPRS Symposium, International Archives of Photogrammetry and Remote Sensing*.

Haris, K., Efstratiadis, S. N., Maglaveras, N., and Katsaggelos, A. K. (1998). Hybrid image segmentation using watersheds and fast region merging. *IEEE TIP*, 7(12).

He, X. and Yung, N. (2004). Curvature scale space corner detector with adaptive threshold and dynamic region of support. In *ICPR*.

Hermann, S. and Klette, R. (2005). Global curvature estimation for corner detection. In *Image Vision Computing New Zealand*.

Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *PAMI*, 20(11).

Jianbo, S. and Jitendra, M. (2000). Normalized cuts and image segmentation. *PAMI*, 22(8).

Juan, O., Kerivan, K., and Postelnicu, G. (2006). Stochastic motion and the level set method in computer vision: Stochastic active contours. *IJCV*, 69(1).

Levin, A. and Weiss, Y. (2006). Learning to combine bottom-up and top-down segmentation. In *ECCV*.

Liu, H.-C. and Srinath, M. (1990). Corner detection from chain code. *Pattern Recognition*, 23.

Mould, D. (2003). A stained glass image filter. In *14th Eurographics Workshop on Rendering*.

Olmos, A. and Kingdom, F. (2006). Automatic non-photorealistic rendering through soft-shading removal: a colour-vision approach. In *ICVVG*.

Perona, P. and Malik, J. (1991). Scale-space and edge detection using anisotropic diffusion. *PAMI*, 12(7).

Reche, P., Urdiales, C., Bandera, A., Trazegnies, C., and Sandoval, F. (2002). Corner detection by means of contour local vectors. *Electronics Letters*, 38.

Rother, C., Kolmogorov, V., and Blake, A. (2004). Grabcut: Interactive foreground extraction using iterated graph cuts. In *SIGRAPH*.

Shotton, J., Winn, J., Rother, C., and Criminisi, A. (2006). Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In *ECCV*.

Sun, J., Jia, J., Tang, C.-K., and Shum, H.-Y. (2004). Poisson matting. In *SIGGRAPH*.

Tomaz, F., Candeias, T., and Shahbazkia, H. (2003). Improved automatic skin detection in color images. In *Digital Image Computing: Techniques and Applications*.

Tran, T. T. H. and Lux, A. (2004). A method for ridge extraction. In *ACCV*.

Vezhnevets, V., Sazonov, V., and Andreeva, A. (2003). A survey on pixel-based skin color detection techniques. In *Graphicon*.

Viola, P. and Jones, M. (2001). Robust real-time object detection. In *CVPR*.

Winn, J. and Jojic, N. (2005). Locus: Learning object classes with unsupervised segmentation. In *ICCV*.

Winnemöller, H., Olsen, S. C., and Gooch, B. (2006). Real-time video abstraction. In *SIGGRAPH*.

Yang, L., Meer, P., and Foran, D. (2007). Multiple class segmentation using a unified framework over mean-shift patches. In *CVPR*.

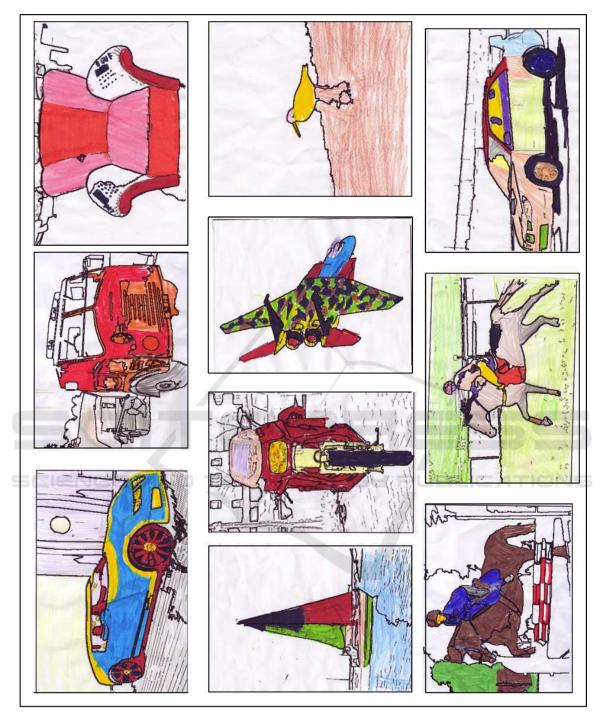Yang, M.-H., Kriegman, D., and Ahuja, N. (2002). Detecting faces in images: A survey. *PAMI*, 24(1).

Figure 10: Example of coloring pages colored by Anton (11 years), Gabriel (9 year), Johanna (8 years), Elisabeth (7 years) and Mikhaël (5 years).