J. Vis. Commun. Image R. 36 (2016) 149–158

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Cartoon filter via adaptive abstraction ☆

Davide Guastella, Cesare Valenti *

Dipartimento di Matematica e Informatica, Università degli Studi di Palermo, Italy

ARTICLE INFO

Article history:
Received 2 February 2015
Accepted 25 January 2016
Available online 1 February 2016

Keywords:
Multiresolution abstraction
Cartoon filter
Redundant wavelets
Circular median filter
Fast multi-scale median
Mathematical morphology
Euclidean distance transform
Edge preserving smoothing

ABSTRACT

Abstraction in computer graphics defines a procedure that discriminates the essential information that is worth keeping. Usually details, that correspond to higher frequency components, allow to distinguish otherwise similar images. Vice versa, low frequencies are related to the main information, which are larger structures. Contours themselves may also be identified by high frequencies and separate each pictured component. The underlying idea of the proposed algorithm consists in identifying these edges, by a redundant wavelet transform, and in blurring the inner areas of the components, by an adaptive circular median filter. In spite of its implementation simplicity, our unsupervised methodology provides results similar to those obtained by more complex techniques already described in the literature.

1. Introduction

Abstraction algorithms return non-realistic scenes, still keeping their main information content. These techniques are widely used in computer graphics with applications in the videogame (e.g. "Prince of Persia", 2008) and in the film (e.g. "A scanner darkly", 2006) industries. In the former case, the so-called cel-shading technique [1] reduces in real time the amount of data that describe the three-dimensional scene with a high quality level. The graphic tool used to model the objects let's the user arrange position, texture and properties of the virtual environment, including lighting and the camera location. This allows to quickly simulate a cartoon-like effect without the actual need for a hard intervention by an artist. In the former case, the interpolated rotoscoping techniques [2] return frames which show a typical effect due to handmade drawings.

Most abstraction algorithms apply a variety of methods to reduce the amount of details within the areas that correspond to the objects in the image, while maintaining the salient information in proximity of their more highlighted contours. Interesting reviews of the main stylization methods can be found in [3,4].

Segmentation-based techniques are usually avoided due to over segmentation and require a continuous supervision to mitigate eventual errors [3]. Anyhow, these segmentation-based approaches are unsuitable on video streams, because of their excessive computational time. Segmentation techniques based on clustering, like superpixel [5] or normalized cuts [6] were not taken into account because of their dependency on minimization functions that involve user-defined parameters (e.g. the cluster size or the affinity function) which depend on the given image and therefore are unsuitable on video streams.

An iterative technique based on the minimization of the gradient information is presented in [7] to globally maintain and possibly enhance the most prominent set of edges by increasing steepness of transition while not affecting the overall acutance. An approximate solution is achieved, because finding the global optimum is a NP-hard problem.

The bilateral filter [8] is based on blurring through appropriate weighted Gaussian functions that preserve as much as possible the areas close to the edges. Again, the main drawback is the computational complexity caused by the non-separability of the proposed functions [3] although recent implementations result quite fast [9]. A framework of different methods was described in [10] where the bilateral filter is applied to further smooth the interior of uniform regions. The subsequent difference of Gaussians enhances the contours and a particular quantization technique reduces the color space palette of the final image.

A pipeline for image abstraction that exploits color quantization, bilateral filtering and difference of Gaussians was introduced in [3,11]. The bilateral filtering was modified conveniently to set the eccentricity of the smoothing window according to the local

☆ This paper has been recommended for acceptance by Yehoshua Zeevi.

* Corresponding author at: Dipartimento di Matematica e Informatica, via Archirafi 34, 90123 Palermo, Italy.

E-mail address: cesare.valenti@unipa.it (C. Valenti).

http://dx.doi.org/10.1016/j.jvcir.2016.01.012
1047-3203/

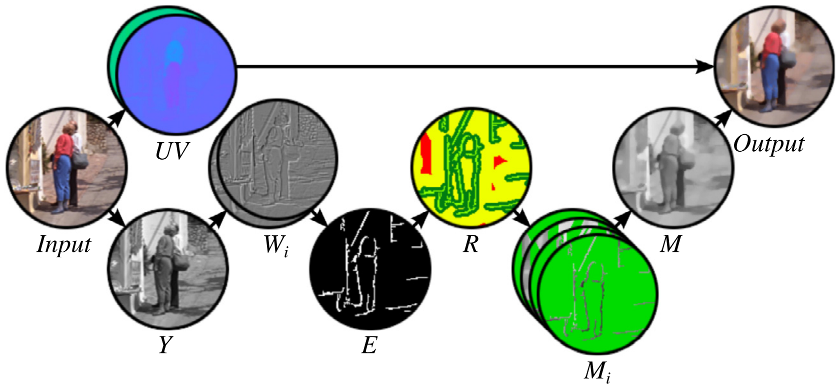

Fig. 1. Sketch of the main steps of our methodology.

gradient. Also, a particular thresholding function was employed to perform color quantization.

The method developed by Litwinowicz [12] is widely used in the cinematographic field. It computes the optical flow between two consecutive frames to determine the main direction to orient a sequence of simple brush strokes. Input parameters like thickness, length, hardness and color of the brush require an accurate control by the artist and they greatly influence the quality of the abstract representation.

Implementations based on the Kuwahara filter [13,14] divide the image into partially overlapped square blocks and assign values to the their central pixels according to the average brightness of the blocks, weighted by a Gaussian convolution [15]. These approaches are generally fast but return images that suffer due to block artifacts. The variant presented in [16] considers circular sectors instead of square blocks and arranges the eccentricity [13] according to an anisotropy function calculated through the eigenvalues of the Harris matrix [17]. This increased complexity makes it possible to get better results even with fixed size windows. A multi-scale approach which uses a pyramid to guide both the shape and the size of the windows was reported in [18] to reduce details with different dimensions.

This paper describes a simple yet effective abstraction methodology made by a pipeline that can be computed in a fast way (see Fig. 1). The following sections describe the main steps which are luminance channel extraction, edge detection by the 'à trous' wavelet, distance transform by mathematical morphology and adaptive smoothing filter. A comparison with state-of-the-art algorithms is reported together with conclusions and future works.

2. Luminance conversion

In order to obtain a valid representation of the luminance information, we compared various techniques as suggested in [19]. Indeed, we verified experimentally that all these methods are qualitatively equivalent to each other for the final abstraction (Fig. 8). Therefore, we decided to convert color to grayscale images by extracting the standard luminance information. Given the RGB color representation of a pixel $\mathbf{p}$, its YUV counterpart is obtained by:

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix}_{\mathbf{p}} = \begin{bmatrix} 0.30 & 0.59 & 0.11 \\ -0.15 & -0.29 & 0.44 \\ 0.62 & -0.52 & -0.10 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}_{\mathbf{p}} \tag{1}$$

where Y takes into account the non-uniform human perception of primary colors (Fig. 2).

3. Edges by wavelets

The wavelet analysis is a powerful mathematical tool for representing and processing data, to enhance or to suppress components with specific frequencies (i.e. size and shape). We applied the so-called *à trous* algorithm [20] because it is very fast and retains the maximum resolution (i.e. the output image does not undergo decimation unlike the usual multiresolution analysis [21]).

We perform a *filterbank* (i.e. succession of low-pass and high-pass filters) on the luminance Y channel:

$$I_0(\mathbf{p}) = Y(\mathbf{p}), \quad I_i(\mathbf{p}) = I_{i-1}(\mathbf{p}) \circledast \ell_i \tag{2}$$

where the non-zero elements of the low-pass filter ℓ_i are given by the isotropic kernel ℓ [22–24]:

$$\ell = \frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}, \quad \ell_i(2^{i-1}\mathbf{q}) = \ell(\mathbf{q}) \tag{3}$$

and the pixel $\mathbf{q}$ spans the 3×3 neighborhood centered in $\mathbf{p}$. The high-pass filter is defined as the difference between two consecutive spatial scales, which provides the wavelet coefficients:

$$W_i(\mathbf{p}) = I_{i-1}(\mathbf{p}) - I_i(\mathbf{p}). \tag{4}$$

This wavelet algorithm takes a constant time when computing a series of W_i due to the advantage that the number of non-zero elements in ℓ_i is always equal to nine (Fig. 3) and the convolution can be speeded up by considering the separability property of ℓ:

$$\ell = \frac{1}{16}\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}[1 \quad 2 \quad 1]. \tag{5}$$

Small objects are enhanced in the first planes W_1 and W_2. A threshold is applied to binarize these planes and the set of contours E is obtained by a subsequent pixel-wise logical disjunction of the binarized images:

$$E(\mathbf{p}) = \bigvee_{i=1,2} \{W_i(\mathbf{p}) > \mu_i - 1.5 \times \sigma_i\} \tag{6}$$

where μ_i and σ_i are respectively the mean and standard deviation of the floating-point values of the plane W_i. Sometimes, the resulting edge includes very small components caused by noise or negligible details in the input image. The usual mathematical morphology opening φ with structuring element S defined by the discrete disk of radius equal to one pixel improves the final result (Fig. 4) [25]:

$$E^{\varphi}(\mathbf{p}) = \delta(\varepsilon(E(\mathbf{p}))) \tag{7}$$