Automatic Trimap Generation with Structured Lighting and Chromakey

Antonio C. T. Blaia Junior, Daniel Z. Vielmas, Henrique F. de Arruda, Ivan A. Aguilar, Antonio C. Sementille
Departamento de Ciências da Computação,
Universidade Estadual Paulista – UNESP,
Bauru, Brazil.
e-mail: jrblaia@yahoo.com.br, danielvielmas@fc.unesp.br, h.f.arruda@gmail.com, ivan_aguilar@fc.unesp.br, semente@fc.unesp.br

(a) (b) (c) (d) (e)

Figure 1: (a) RGB picture. (b) Depth map. (c) Trimap – foreground (red), background (black), "unknown" (yellow). (d) Alpha map. (e) Composite image.

(a) (b) (c) (d)

Figure 2: (a) RGB picture. (b) Trimap – foreground (red), background (black), "unknown" (yellow). (c) Alpha map. (d) Composite image.

***Abstract* — This project has as an objective of the extraction of humans in the foreground of image by creating a trimap which combines a depth map analysis and the Chromakey technique. The trimap is generated automatically, differing from the manual implementations which require user interaction. The extraction is based on extra information deriving from a structured lighting device (Kinect) integrated with a high resolution camera. With the junction of the monochromatic Kinect camera and the high definition camera, the results so far have been more expressive than only using the RGB and monochromatic cameras from the Kinect.**

Keywords: Matting, Chromakey, Trimap

I. Introduction

The images displayed on video, broadcast television or movie scenes can be imagined as a composite of two layers: one containing foreground objects and one containing the background.

Both the television and film industry have the resources to use and isolate the foreground layer, creating the need of removal from its original context. This separation, known as key extraction, aims at a future combination among layers of backgrounds originated from several sources that can be done by filming in other locations or images created on computers. Several studies in the key extraction are protected by patents [1] [2].

Digital matting is a process of composing digital images, a foreground element is extracted from a background image by estimating a color and opacity for the foreground element at each pixel [3].

In order to identify the objects to be removed from the background, a three-level pixel map can be created by the user which segments the input image into three regions: foreground, background and unknown. This map is often referred as trimap [3].

Contributions: This project automates the digital image matting, thus aiding in the low-cost film productions, serving as a tool for a virtual studio.

II. Metodology

Initially the project was developed for the use of a structured light device, the Kinect, this device has an infrared light emitter, a RGB camera and a monochrome camera sensitive to infrared.

These components allows the device to track human beings, with this extra functionality as information we use the depth map along with the human being highlighted, as shown in the Figure 1b.

Given the approximation of the monochrome image to RGB and the area of uncertainty around the generated object by the device itself contributed to the presence of noise in the depth map. Thus the solution was the application of two filters: dilation and Gaussian.

An analysis is made on all the pixels, each image point is studied along with its corresponding point on the depth map. Knowing the color information of the pixel and its depth, comparisons are made, which consider the background color (green or blue) and depth. This method generates a trimap, predicting the areas of certainty and uncertainty (Figure 3).

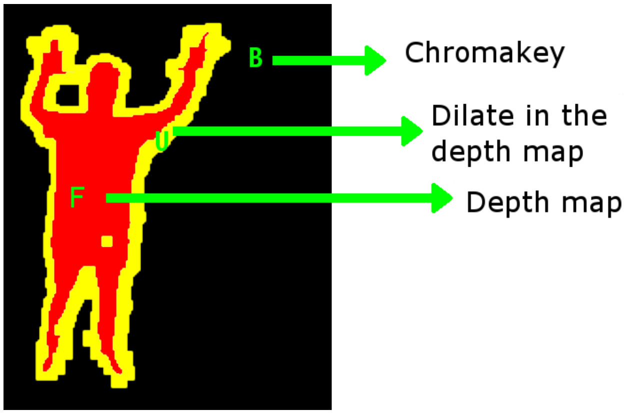

Figure 3: Trimap generated by depth map

Aiming more significant results, we use a high resolution camera (FullHD) as a source image acquisition.

Instead of using a single range of colors to identify the arbitrary color (green or blue), two ranges of colors are used by setting two different values for the Euclidian distance [4]. Pixels with color within the first range are classified as background, while pixels with color within the second range are classified as an unknown pixel. Pixels excluded from these two ranges are considered foreground (Figure 4).

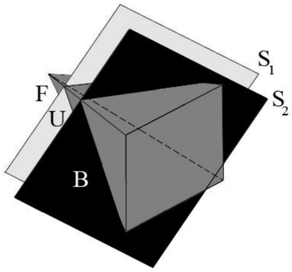

Figure 4: An RGB cube showing two planes S1 and S2 intersecting the skew pyramid

With the high resolution camera there was a gain in trimap precision, consequently the results of digital matting algorithms were better than those obtained by the Kinect.

The gain in precision in the generation of trimap with the high resolution camera, motivated us to engage the high resolution camera to the Kinect, replacing the low-resolution RGB camera on the Kinect, taking advantage of the extra information of the depth map with the Kinect RGB image with the high resolution camera (Figure 4).

This junction in devices created the need to establish the correspondence between the pixel and the depth map image pixel RGB from the high resolution camera. So we calibrated the cameras with a chessboard where the corners of the squares are marked for further correspondence between the images. Thus we find the transformation matrix to correlate the pixels.

III. EXPERIMENTATION

Two sets of tests were conducted. The first set used both of the Kinect's cameras (RGB and monochrome) applying Chromakey techniques and the extra information provided by the depth map (Figure 1). The second set of tests used only the high resolution camera to apply the Chromakey technique (Figure 2).

The tests from the second set were performed with the actor at a distance of 90cm from the high resolution camera and 100cm from the wall.

The current phase of the project is to generate new tests from the newly implemented transformation matrix.

IV. CONCLUSION AND FUTURE WORK

The automatic generations of trimaps, along with the use of a high resolution camera coupled to the Kinect are significant contributions of the project.

With the transformation matrix it will be possible to make the junction between the developed techniques and make a comparison between the results obtained so far.

Performance analysis will be made between the methods, so that this way it will be possible to use in specific applications, as well as studying the feasibility of applying the same in video.

For future work, there is the possibility of using massively parallel processing, using graphics cards, in the automatic generation of the trimap and algorithms in digital matting.

REFERENCES

[1] VLAHOS, P. Composite photography utilizing sodium vapor illumination. United States Patent n. 3,095,304 25 jun. 1963.

[2] VLAHOS, P. Eletronic composite photography. United States Patent n. 3,595,987 27jul. 1971.

[3] Y.-Y. Chuang, B. Curless, D. H. Salesin, and R. Szeliski, "A bayesian approach to digital matting," in Proceedings of IEEE CVPR, pp. 264–271, 2001.

[4] BERGH, V. D.; LALIOTI, V. Software chroma keying in an immersive virtual environment. South African Computer Journal, v. 24, p. 155–162, 1999.