AUTOMATIC BODY ANALYSIS FOR MIXED REALITY APPLICATIONS

D. DOUXCHAMPS, X. MARICHAL, T. UMEDA

Laboratoire de Télécommunications et Télédétection - UCL
Bâtiment Stévin, 2 place du Levant, B-1348 Louvain-la-Neuve, Belgium
E-mail: {*d.douxchamps,umeda ,marichal}@tele.ucl.ac.be*

P. CORREA HERNANDEZ*, F. MARQUES

Department of Signal Theory and Communications - UPC
Campus Nord, Modul D5, C/ Jordi Girona 1-3, 08034 Barcelona, Spain
E-mail: {pedro,ferran}@gps.tsc.upc.es

The present paper elaborates on specific techniques to improve the performance of a technical system that was designed and implemented by the IST-*art.live* project to offer multimedia authors the capacity of designing immersive interactive narratives involving real people into their own universe of pictures, graphics and associated designs. The resulting mixed-reality environments allow for the creation of stories that mix graphical elements with inputs from live cameras. When some users come in front of the cameras, they get themselves immersed within the visual ambiance and they are therefore involved within the narrative, which they are able to interact with through their gestures and behavior. The present paper introduces new results about the body analysis of the users and the interactive capabilities it provides: head, hands and feet of the users are detected and tracked thanks to geodesic distances with respect to the center of gravity of the segmented silhouettes and human morphological information.

1. The "Transfiction" Concept

Overall, by Transfiction, we mean "**trans**portation in **fiction**al spaces" as illustrated on figure 1. Such transportation occurs thanks to immersion and interaction.

In the framework of Mixed Reality [1], Transfiction [2] is designed for mixing synthetic and natural images in real time and allows one to interact in these input/output screens through natural gesture and speech. Transfiction systems are intended for intuitive interaction in a non-obtrusive manner,

* This work is supported by the Walloon Region under FIRST EUROPE Objectif 3 grant EPH3310300R0132/215069 and the grant CICYT TIC2001-0996 of the Spanish Government.

allowing one to develop a novel media and to offer users rich and engaging experiences.

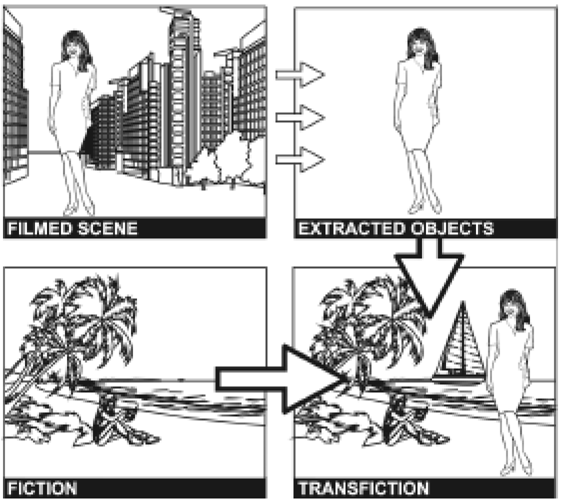

Figure 1: Transfiction concept.

2. System Implementation

Figure 2: Typical Transfiction situations (background image on the left is © Casterman-Schuiten).

One of the peculiarities of the Transfiction system is the willingness to have it performing in real-time on standard PC architecture. Real-time is indeed mandatory in terms of design for users to enjoy the experience in settings using the magic mirror paradigm (cf. figure 2): achieved framerate after all treatments (from acquisition to rendering, including analysis) must be superior to 15 fps; latency must be kept inferior to 1/25 s. The choice of a PC architecture allows developers to benefit from the many advances of graphic cards, bus and CPU speeds while ensuring a maximum availability of the research results for concrete applications. Therefore, all investigated techniques for body analysis must be able to perform faster than real-time on such architecture, while being robust to noise and segmentation artifacts.

3. Body Analysis

In order to extract the exact 3D posture of the human subject we use three orthogonal cameras. This allows to get rid of self-occlusions, and to obtain the positions of, at most, five crucial points (see further) in the three different views. For each view, the silhouette of the actor is extracted using a simple, real-time segmentation technique [3] that it is applied on a controlled scene. The analysis algorithm then has two main steps:

Selection of crucial points: crucial points are defined as the five most prominent human features in the silhouette representation. This notion of prominence can be translated in terms of distance from the center of gravity (*CoG*) to the silhouette. Typically, the extraction of human features is achieved either using heuristic methods dealing with a priori average human limb lengths [4], or with non heuristic methods which extract the maximum possible amount of information from the human silhouette [5].

The second approach has been selected: a robust method is developed to search for crucial points. It relies on the extraction of points of the silhouette that represent local maxima of the geodesic distance with respect to the CoG: one has to compute the geodesic distance map of the silhouette with respect to the CoG. By contour tracking the silhouette, a one-dimensional function is computed representing the geodesic distance function of the silhouette points with respect to the CoG. This function presents local maxima associated to the crucial points as well as other noisy local maxima. The latest ones are removed by eroding the function. The erosion factor can be chosen a priori, knowing the noise characteristics of the cameras and the image size.

Labeling of crucial points: For this purpose, some human morphological information is needed. At first, the detected crucial points are used to build a robust skeleton of the actor, only containing branches connected with those crucial points (cf. figure 3). This skeleton allows for the estimation of the points that are geodesically closest or farthest to the rest: five geodesic distance maps of the skeleton are computed, one for each crucial point. The distances of each crucial point to the one generating the distance map are added up (*scores*). The two crucial points having the largest scores are labeled as being the feet. The other three points represent the group of hands and head. The point labeled as being the head is the point of the previous group that has the more symmetrical distance with respect to the other two (cf. figure 3).

This algorithm has been adapted to handle the situations where less than five crucial points have been detected (some crucial points being omitted by self occlusions). The labeling phase ends with a verification routine that checks whether some morphological aberrations do not occur (e.g. that the feet are not above the head). The same algorithm is applied to the front and side view. A slightly modified version is applied to the upper view.

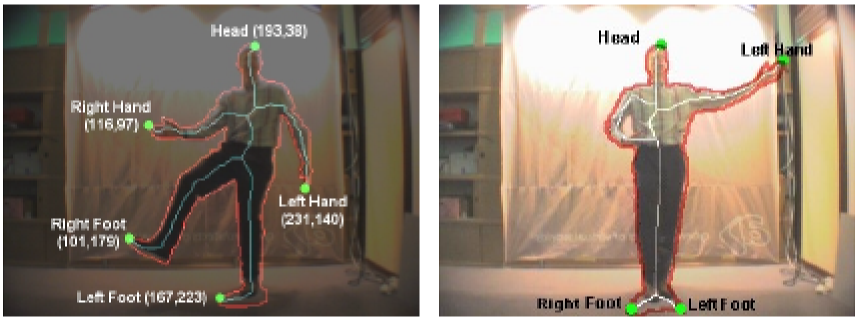

Figure 3: Labeled crucial points and final skeleton on two frontal views

References

1. P. Milgram and H. Colquhoun, *A Taxonomy of Real and Virtual World Display Integration*, in *Mixed Reality, Merging Real and Virtual Worlds*, Y. Ohta and H. Tamura editors, Ohmsha-Springer (1999)
2. A. Nandi and X. Marichal, *Transfiction*, chapter 2, pp. 13-33 in *CyberText - Yearbook 2000*, M. Eskelinen and R. Koskimaa editors, Publications of the Research Centre for Contemporary Culture 68 (2001) ISBN 951-39-0905-0
3. The art.live consortium, *The ART.LIVE architecture for Mixed Reality*, Virtual Reality International Conferences, Laval, France (June 2002)
4. J. Ohya, A. Utsumi, and J. Yamato, *Analyzing Video Sequences of Multiple Humans: Tracking, Posture Estimation and Behavior Recognition*, The Kluwer International Series in Video Computing 3 (March 2002) ISBN 1-4020-7021-7
5. P. Soille, *Morphological Image Analysis, Principles and Applications*, Second Edition, Springer-Verlag (2003)