Matching for Perceptual User Interface

Jose Maria Buades Rubio
Universitat Illes Balears
Ed. Anselm Turmeda
CrtaValldemossa Km 7.4
Spain (E-07122), Palma Mca
josemaria.buades@uib.es

Francisco Perales López
Universitat Illes Balears
Ed. Anselm Turmeda
CrtaValldemossa Km 7.4
Spain (E-07122), Palma Mca
paco.perales@uib.es

Javier Varona Gómez
Universitat Illes Balears
Ed. Anselm Turmeda
CrtaValldemossa Km 7.4
Spain (E-07122), Palma Mca
xavi.varona@uib.es

ABSTRACT

In this paper we describe a complete method for building a perceptual user interface in indoor uncontrolled environments. Overall system uses two calibrated cameras and does initialization: detects user, takes his/her measurements, builds a 3D-Model; and also performs matching/tracking for: trunk, head, left arm, right arm and hands.

Keywords

Human Computer Interaction, Perceptual User Interface, Matching, Virtual Reality Interaction

1. INTRODUCTION

Overall system uses two calibrated cameras and does initialization: detects user, takes his/her measurements, builds a 3D-Model; and also performs matching/tracking for: trunk, head, left arm, right arm and hands. System is waiting for a user in a predefined posture, once user has been detected he/she is analysed to take measurements and build a 3D-Model. Tracking is carried out by a Montecarlo probabilistic method and divided in three steps, track trunk and head, left arm and right arm, this divide and conquer solution proposed improve computation time without getting worse results. Matching process uses two sub-matching functions, one to compute colour seemed and another to compute shape one.

2. INITIALIZATION

Initialization is performed from one camera. After user appears in action system detects him [Bua03] and takes measurements for later matching process, as shown in figure 1, system models trunk region as a rectangle, hands, arms and head regions as 3D-ellipsoids. 3D reconstruction is performed taking into account that user is at same distance that calibration object was at calibration process.

Posters proceedings ISBN 80-86943-04-6
WSCG'2006, January 30-February 3, 2006
Plzen, Czech Republic.

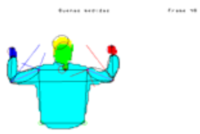

Figure 1. User modeling from single image.

3. MATCHING

The matching process starts in the first frame after user detection, and it is performed taking account the following visual cues: color and shape. Besides, we will use spatial constraints.

Tracking is performed using a Montecarlo strategy in two steps. First, the system try to match the trunk and head segments. After, the arms (left and right separately) are matched using the previously computed trunk position. These two steps reduce the computing time, instead of perform a global search, we benefits from divide and conquer strategy.

Search Strategy

As we have commented before, we exploit the benefits of a divide and conquer strategy. The trunk segment area is the large region and never is occluded. For this reason, this strategy obtains good results; else the results would be unexpected positions when an occlusion occurs. Second step is to perform matching for the most interesting body parts, the hands. For each arm is performed a Montecarlo search in an independent way. This three matching process (trunk and head – left arm – right arm) reduces computation, instead of multiply computation time for trunk, left and right arm, computation times are added. The results are quite good as it is in the results section. Proposed positions are evaluated as biometrically possible to avoid undesirable results and reduce computation time.

Matching Function

The matching function is composed of two parts: colour comparison and shape comparison.

The trunk segment is modelled as a 2D rectangle, others shapes as a 3D box have been tested but surprising the results are worse. The rest of segments (arm, forearm, hand and head) are modelled using super-ellipsoids.

For each segment, the colour model is computed from initialization process (a better solution would be to have a texture model). The segment is projected in the image to evaluate a colour matching function; each pixel is compared with the colour model and scored as good (1) or bad pixel (0), for trunk segment only are projected a grid of 15x10 pixels with so good results as projecting all the segment. The other segments, super-ellipsoids, also a sample of pixels are projected. For each segment colour matching function returns the ratio of good pixels detected.

The shape matching function uses the contours. To detect contour in the captured image performs Sobel operand, as a result we get an image, called Sobel-image. Segment contours are projected in Sobel-image and scored pixels as good (1) or bad (0) in relation with its value. This function returns the percentage of pixels scored as good. This function has good results if the segment is just in the contour of the image, and it is very difficult because the real person shape is very deformable. A better solution is to take into account the distance from the segment contour to the contour in the Sobel-image. In this second solution not only the exact pixel is compared, neighbour pixel in normal direction (from near to far) are evaluated, thus a pixel is evaluated as a value between 0 and 1, in relation to the distance to a contour in sobel-image, more near a greater value.

Matching function returns a value for a segment s_j as follows:

$$m(s_j) = \min_{\forall \gamma_i \in H}(m^{\gamma_i}_{colour}(s_j) + m^{\gamma_i}_{shape}(s_j)) \quad (1)$$

where H is a set of cameras, $\{\gamma_1,\gamma_2\}$ in this case, *s* is a segment, $m^{\gamma_i}_{colour}(s)$ is colour matching function from camera γ_i, $m^{\gamma_i}_{shape}(s)$ is shape matching function.

Finally, the Matching value for a pose Ψ is defined in Equation (2).

$$m(\Psi) = \sum_{j=1}^{N} m(s_j) \quad (2)$$

4. ACKNOWLEDGMENTS

This work has been subsidized by Universitat Illes Balears, and the national project TIN2004-07926 from the MCYT Spanish Government. Acknowledges the support of a Ramon y Cajal fellowship from the Spanish MEC.

5. REFERENCES

[Bal96] C. Ballester, V. Caselles and M. Gonzalez, “Affine invariant segmentation by variational methods”, SIAM J. Appl. Math., Vol. 56, No 1, pp. 294-325, 1996

[Har00] I. Haritaoglu, “W4: Real-Time Surveillance of People and Their Activities” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 22 No8, pp 809-830, 2000

[Sid00] H. Sidenbladh, M.J. Black and D.J. Fleet “Stochastic Tracking of 3D Human Figures Using 2D Image Motion” ECCV 2000.

[Bua03] J.M. Buades, M. Gonzalez, F.J. Perales. “A New Method for Detection and Initial Pose Estimation based on Mumford-Shah Segmentation Functional”. IbPRIA 2003. Port d’Andratx. Spain. June 2003. pp 117-125

Figure 2. Tracking results.