Human motion analysis and synthesis using graphical biomechanics models applied to disable swimming people

A. Aguiló*, P. Martinez*, J.M. Buades, F.J. Perales**, M. Gonzalez****

**Nursing and Physical Therapy Department*

***Computer Graphics and Vision Group - Department of Computer Science*

Universitat de les Illes Balears (UIB) – Spain

Telephone: +34 971 17 30 00, Fax: +34 971 17 30 03

E-mail: paco.perales@uib.es

Abstract

In many applications the study of human movement using a computer vision and graphics techniques is very useful. One of these applications is the three-dimensional reconstruction of the structure of the human body and its movement using sequences of images and biomechanical graphics models. We present a special study about disable people and swimming activities, high level competition in particular. This kind of study needs accuracy on the analysis and reconstruction of the person's body, therefore the virtual human (avatar) must have similar anthropometric characteristics than the person who is doing the movement. We define a semi-automatic process to adjust the humanoid to the morphology of the person. It could be very laborious and subjective if done manually or by selection of points, but in this article we present a global human motion system capturing, modeling and matching a semiautomatic process between the real person and the modeled humanoid or synthetic avatar.

Once this process is carried out we are ready to analyze and represent the movements under study. It must be defined a specific set of parameter for every kind of sport. The study is adapted to specific sport activities, in our case swimming activities with an additional complexity: water occlusion and distortion. In these cases the adjustment process must be assisted by the computer using rules and models to help the expert user in correspondence tasks. Until now it is been used in indoor spaces and controlled environments. The system requires no markers or special clothing to be worn by the swimmer, and therefore its range of application is very wide. In other cases, special clothing could help very much in segmentation process. It is also very important its portability into domestic environments using VRML 2.0 and the H-anim standard for specification of virtual humanoids. In the following sections we explain the whole system, the capture process, the segmentation process, some of our results and future work.

Keywords

Humanoid, real and synthetic Images, VRML, segmentation techniques, colour models, functional segmentation, matching, calibration, graphic model, sports motion, handicapped performance, rehabilitation evaluation.

1. Introduction

The problem of modeling, tracking and 3d reconstruction of human body is a central task of computer vision and graphics community. At this time, a variety of human bodies modeling techniques are available, that can be classified into the creative approach and reconstructive area. We are more oriented to the second one; much work has been devoted to this reconstructive approach to build 3D geometry of human body in automatic way. As key examples we can suggest the work done in [11, 12, 13] in these papers the authors use 3D range scanner and PCA but we plan to use digital cameras configuration with less accuracy but low cost. Also another's examples are showed in the research papers in [14, 15, 16]. In these research work the use of cameras and computer vision tracking techniques are the kernel of systems.

Our proposed, system is oriented on the before ideas proposed. That's means that multiple cameras and computer vision techniques are the kernel. Also, non-invasive devices and real time constrains are fundamental to apply our system to virtual and augmented reality systems.

We will divide the general process of the system into four stages: in the first one we capture images of the person from different points of view and of the background in particular up to four IEEE 1934 color cameras are used, that means that a initialization process is needed to know the exact anthropometrical data of the person that is moving in front of the cameras. In the second stage, we select the humanoid with similar characteristics to the original individual, so the result of this initialization process is an avatar with the same segments length of the real person. This process is need because in high-level sport activities the accuracy is very critical. In the following stage we apply a semiautomatic process with which we obtain the humanoid adjusted to the person's measurements. The final goal is to reach an automatic recognition process, but this is a challenging task that we are solving for less accuracy tasks and under controlled environments. In the case of sport activities, a semiautomatic system is a good trade-off solution between full automatic systems and commercial standard systems using markers. The last stage combines the captured images and the generated humanoid to verify the result of the process and study the values we are interested in. We can use numerical data to make a deep study of performance of movements or paint special lines or points in real and synthetic images to help the instructor in the training process to reach the perfect performance in sports for handicapped people.

2. Capturing process, Modeling & Standards Humanoids

The capture software has been implemented with the API designed by Microsoft for Windows platform, this API called DirectShow allows you to use any camera (IEEE 1394, USB Web Cam, parallel port scanner, video file…) as long as you have drivers for Windows. Any kind of these input devices is programmed in a transparently and independent hardware way, without necessity to modify our application. This API has been chosen with the intention to cover the most number of end users at a low cost without changing capturing system. So the results of this implementation have the properties: versatility, independence of hardware, oriented object programming, efficiency and easily parallelizable. There are various specifications [2] for humanoids; we have chosen the one created by the group h-anim in VRML format for its portability and adaptability to different applications. The humanoid is composed of a collection of joints and segments structured in the form of a tree. Each joint corresponds to an anatomical joint (knee, shoulder, vertebrae), for example with each vertebra hanging from the one above, and the wrist joint hanging from the elbow joint. Each joint has associated with it a segment it represents, for example the elbow has the upper arm as its segment.

Each segment may have sites and displacers. A site corresponds to an endpoint, such as the fingertips, while a displacer corresponds to an effect applied to the segment. In particular we would like to define a multilevel system. That means that depending on the accuracy of applications, the system could include more degrees of freedom. For human motion recognition task in particular, the model is simpler and for human graphical simulations the system is more complex.

Also, a modelling editor has been developed to adapt the human limitations of handicapped people in this application. A simple animation tool is also included to generate simple kind of movements.

The final expected result is a virtual model of the real person compliant with H-anim standards and, in the case of handicapped swimmers; a special numerical parameter set is considered. All that process will be described in the next sections.

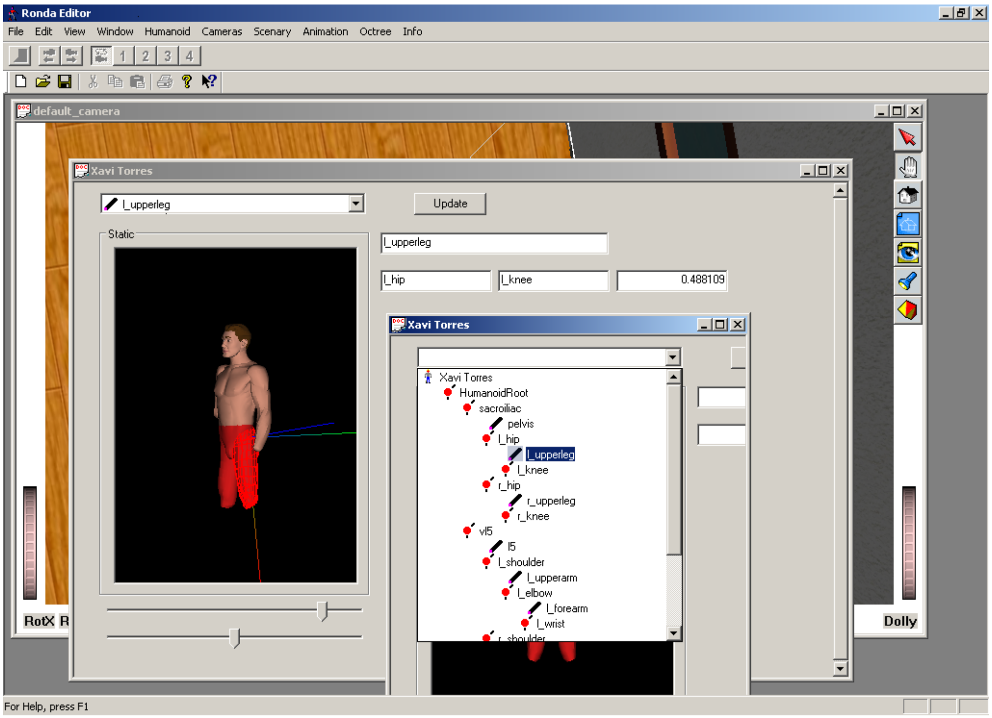

Figure 1: H-Anim biomechanical editor

Another property of the system is the independence between hierarchical structure and graphical representation. That's mean that we can change the graphical primitives and represent the same biomechanical model with wireframe, shapes or volumetric primitives.

3. Segmentation

Image segmentation is the first step in data extraction for computer vision systems. Achieving good segmentation has turned out to be extremely difficult, and it is a complex process. Moreover, it depends on the technique used to detect the uniformity of the characteristics founded between image pixels and to isolate regions of the image that have this uniformity. Multiple techniques have been developed to achieve this goal, such as contour detection, split and merging regions, histogram thresholding, clustering, etc. A survey can be found in [7].

In color image processing, pixel color is usually determined by three values corresponding to R (red), G (green) and B (blue). The distinctive color sets have been employed with different goals, and specific sets have even been designed to be used with specific segmentation techniques [7].

We define a color image as a scalar function $g = (g^1, g^2, g^3)$, defined over image domain $\Omega \subseteq \Re^2$ (normally a rectangle), in such a way that $g\colon \Omega \rightarrow \Re^3$. The image will be defined for three channels, under the hypothesis that they are good indicators of autosimilarity of regions. A segmentation of image g will be a partition of the rectangle in a finite number of regions; each one corresponding to a region of the image where components of g are approximately constant. As we will try to explicitly compute the region boundaries and of course control both their regularity and localization, we will use

the principles established in [8, 10] to define a good segmentation. To achieve our goals we consider the functional defined by Mumford-Shah in [9] (to segment grey level images) which is expressed as:

$$E(u,B) = \int_{\Omega} \sum_{i=1}^{3} (u^i - g^i)^2 \, d\mu + \lambda \ell(B) \qquad (1)$$

where B is the set of boundaries of a homogenous region that define a segmentation and u (each u^k) is a mean value, or more generally a regularized version of g (of each g^k) in the interior of such areas. The scale parameter λ in the functional (1) can be interpreted as a measure of the amount of boundary contained in the final segmentation B: if λ is small, we allow for many boundaries in B, if λ is large we allow for few boundaries.

A segmentation B of a color image g will be a finite set of piecewise affine curves - that is, finite length curves - in such a way that for each set of curves B, we are going to consider the corresponding u to be completely defined because the value of each u^i coordinate over each connected component of $\Omega \setminus B$ is equal to the mean value of g^i in this connected component. Unless stated otherwise, we shall assume that only one u is associated with each B. Therefore, we shall write in this case $E(B)$ instead of $E(u, B)$. A segmentation concept, which is easier to compute, is defined as follows:

Definition 1. *A segmentation B is called 2-normal if, for every pair of neighboring regions O_i y O_j , the new segmentation B' obtained by merging these regions satisfies E(B') > E(B).*

We shall consider only segmentations where the number of regions is finite, in other words $\Omega \setminus B$ has a finite number of connected components and the regions do not have internal boundaries.

A more detailed explanation of the concepts and their mathematical properties can be consulted in [8, 10] and we can see the properties of the functional in [9, 10]. The use of multi-channel images (e.g. color images) can be seen in [10]. We shall use a variation of segmentation algorithm by region merging described in [8] adapted to color images. The concept of 2-normal segmentations synthesizes the concept of optimal segmentation we are looking for, and it lays on the basis of the computational method we use. The whole process is presented in a previous work [4]. For further explanation please refer to this paper. The algorithm uses the RGB components because the segmentations obtained are very near to our goal. But the system is able to use other color space or color descriptor as we can see in [7]. Moreover, if it is needed it can weigh the channels used in order to obtain the segmentation.

After segmentation process we test every region to detect skin regions. Skin color is a characteristic color that is very different to other colours. Skin color test is applied, but underwater skin regions are not detected due to water distortion in light spectrum. In any case the results cases are not bad, and we are using only standard images from commercial video cameras. In the future we plan to use special light conditions and they might be infrared images. Some special clothing at end effectors can also help us in the segmentation process. In the next images we can see some results.

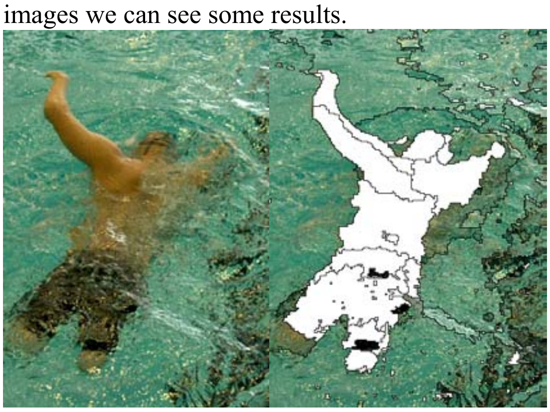

Figure 2: Global body Segmentation

Figure 3: Skin body Segmentation

4. Matching

In the proposed systems the matching process is the kernel. Our main objective is to find a one-to-one correspondence between real and synthetic human body segments and joints in every frame in 3D space. Matching the humanoid is currently done semi-automatically; we are working on automatic processes for a general indoor environment and for the moment we have reached promising results for cases where occlusions aren't so long in time. A trade off solution is a semi-automatic process, that's means, a human interaction in matching process, but supervised by computer that avoids wrong biomechanical postures. The information captured is from the colour cameras, in three positions following specific anthropometric criteria, from which we obtain the desired parameters. We use a database of predefined movements and models to help the matching process. Also we can use this knowledge to estimate new movements that are not previously recorded. The semiautomatic matching process has certain limitations of precision according to the models and the calibration of the cameras. This process is based on a set of well-defined conditions.

The matching process associates each joint with a 2D point in the image. This process consists on analyzing each image obtained from the cameras in an instant of time t. Once we have the articulation located in two or more cameras we estimate the most accurate 3D point; a joint may only be detected in one or none image, therefore the process will have to be completed with contextual information from a higher level. Analyzing the sequence, we are able to apply physical and temporal restrictions, which help us to carry out the matching which reducing the errors and the search space. This adjustment process is conditioned by a set of conditions, which are optimized in each case and type of movement. The restrictions are:

- Angle and distance limitations of the joints
- Temporal continuity of the movement in speed or acceleration
- Prediction of the movement based on a database of known movements, in the case of having non-visible joints.
- Collision of entities.

This set of restrictions defines a level of matching that can follow the whole criteria that's presented in Figure 4. This matching process currently works in a semiautomatic way, thus making that it is possible to work at different levels of precision depending on the application the results are required for.

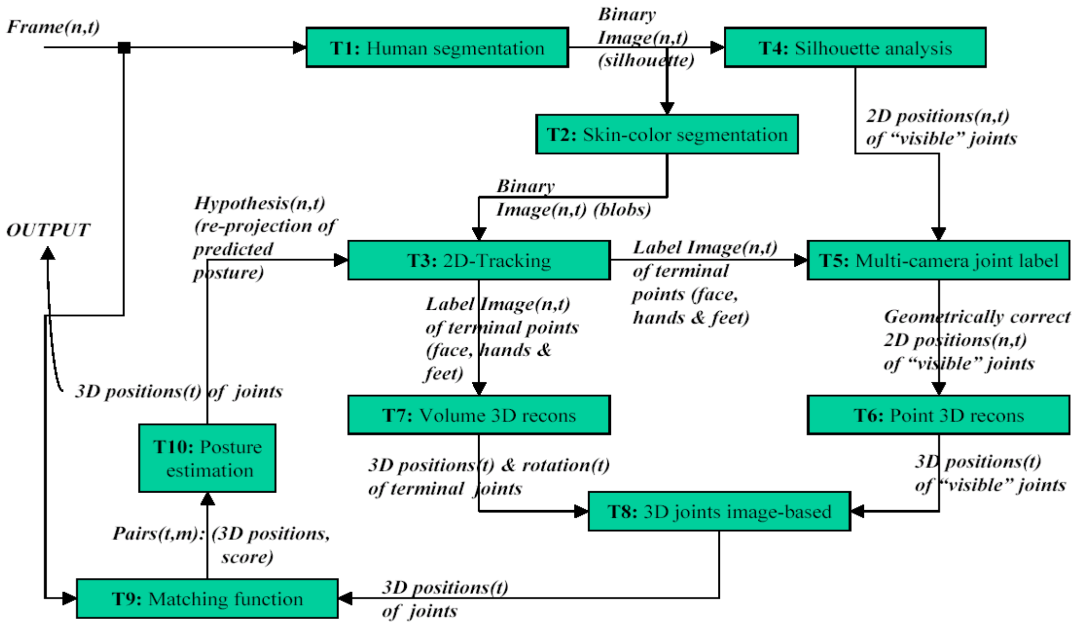

Figure 4: Segmentation, tracking & reconstruction proposed system

In the figures include some matching results can evaluate. In this case the matching is done semi-automatic and without calibration so the results can include some errors. Also the end effectors are not considered (hand, etc.)

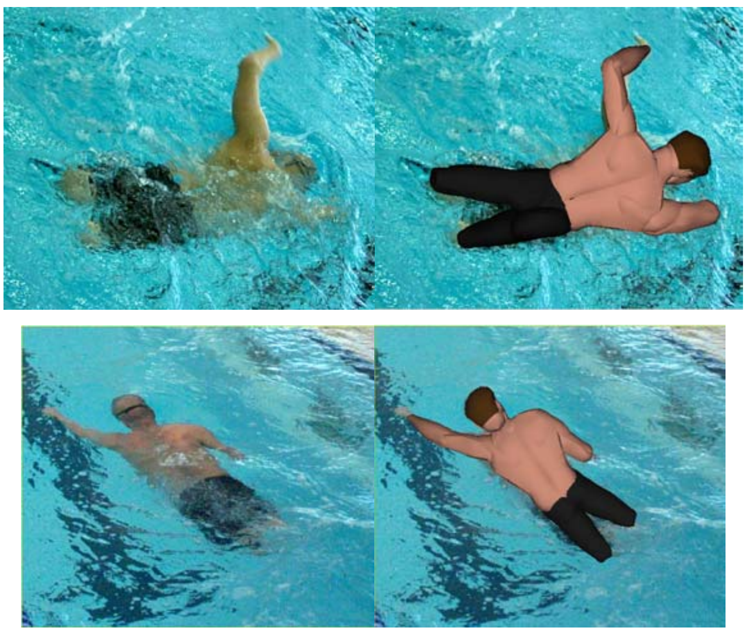

Figure 5: Semiautomatic matching examples with non-calibrated cameras.

5. Biomechanical parameters studied

To study a specific swimming human model we need a set of well-defined parameter. The following list includes the most important ones. The final aim of this work is to obtain all these parameters in an automatic way, but only a subset are reached with computer vision segmentation and analysis techniques for the moment:

- Defining the skew of upper limbs in order to make a difference in the biomechanics affective ness between.
- Calculating and observing the movement of the gravity centre
- Defining the anthropometrics parameters of several segments:
    - Length of both lower & upper limbs
    - Total length of each half-body
    - Distance between both acromioclavicular articulations & distance from each lower limbs to the vertex
    - Length of trunk: distance between the suprasternal point and the middle point between both hips
    - Length of the three segments of the left upper limb:
        - •From the acromioclavicular to the olecranon and from olecranon to the styloid apophysis
        - •From the styloid apophysis to the distal phalanx of the finger

At the moment, all these parameter are computer in a manually or semiautomatic way. Only a subset is computed automatically (for example: centre mass point). In the image 6 we can see a segmentation and centroide computation.

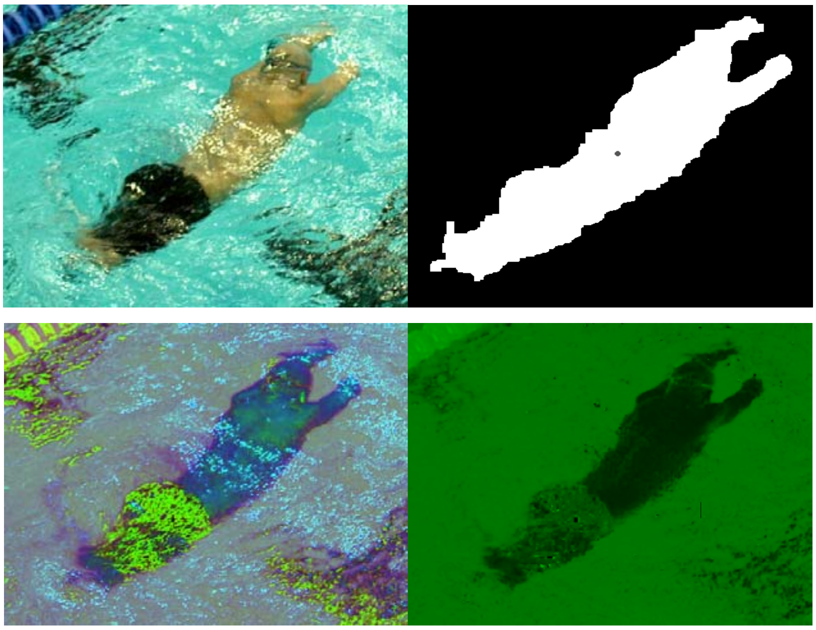

Figure 6: Color segmentation HSL and centroid computation.

6. Conclusions

In this paper we have presented a whole system to analysis and synthesise human movements. The main outstanding research and contributions are: a) Robust programming of a digital IEEE1934 synchronized and portable low cost capturing system (that's isn't a non trivial task!), b) Biomechanical compliant human modelling from our vision oriented model to an H-Anim, c) Robust colour segmentation process using solid mathematical background theory, d) Semiautomatic matching process based on rules from images features and biomechanical conditions e) Good accuracy application with supervised matching criteria in real sport activities and handicapped users. The system isn't really full automatic, but much better than manual systems (in time processing and accuracy). In general the final system should allow us to detect, follow, and recognize the activities of one or more people in a scene [4]. The part presented allows editing the humanoid only under supervision, but a possible extension would be obtaining cinematic and dynamic means with a view to more sophisticated applications. In fact, a determined level of precision will be defined according to the application.

This work is framed within a more general project of analysis and synthesis of human movement using techniques of digital image processing and computer vision. Is very important to remark that the system proposed is non invasive and does not use markers on people. The biomechanical model is also over placed on real images and we don't need individual manual digitalization for every frame and an interactive feedback with the model is possible in real time processing.

The last part presents the adaptation of the proposed system to the analysis of sports motion in all possible variations, in particular for people with several physical limitations. Obviously each discipline has its parameters of interest, which will be defined by the user or specialized technician. The precision of the system is adequate for indoor sequences. We have problems with water interferences in segmentation process and we plan to include underwater images in near future. We are working to improve the segmentation process and reach the biomechanical data as automatically as we can. The manual interaction must be avoided to reach more precision in human matching. We are working with filtering estimation using predictive methods and dynamic and kinetic control of biomechanical solution proposed.

Acknowledgements

This work is subsidized by the European project HUMODAN IST-2001-32202 and Spanish Contract TIC2003-0931 & TIC2002-10743-E. Also we would like to express ours thanks to all members of Computer Graphics and Vision group by their explicit and implicit support to develop this work. Also some video demos can been downloaded at web page http://dmi.uib.es/research/GV/HUMODAN/demonstrators.htm. We also want to make a special mention to Mr. Xavi Torres, a para-olimpic swimmer who has collaborate with us in the experiments http://www.xavitorres.com/

References

1. R. Tsai. "An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision". Proc. IEEE Computer Vision and Pattern Recognition, pp 68-75, 1986.
2. N. Badler, C. Phillips, B. Webber. *Simulating Humans. Computer Graphics Animation and Control*. Oxford University Press, 1993.
3. C. Wren, A. Azarbayejani, T. Darrell, A. Pentland. "Pfinder: Real-Time Tracking of the Human Body". IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp 780-785
4. 4 J.M. Buades, M. Gonzalez, F.J. Perales. "Face and Hands Segmentation in Color Images and Initial Matching" International Workshop on Computer Vision and Image Analysis. Palmas de Gran Canaria. Dec. 2003. pp 43-48
5. Jessica K.Hodgins, Nancy S. Pollard. "Adapting Simulated Behaviors For New Characters". Computer Graphics Proceedings, 1997 pp. 153-162
6. Christopher R. Wren, Alex P. Pentland. "Understanding Purposeful Human Motion", Submitted to ICCV 1999
7. H.D. Cheng, X.H. Jiang, Y. Sun, JinGli Wang "Color Image Segmentation: Advances and Prospects", Journal of Pattern Recognition 34, (2001), pp. 2259-2281
8. G. Koepfler, J.M. Morel, and S. Solimini, "Segmentation by minimizing a functional and the merging methods", SIAM J. on Numerical Analysis, Vol 31, No 1, Feb. 1994
9. D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and variational problems", Communications on Pure and Applied Mathematics, XLII(4), 1989
10. J.M. Morel and S. Solimini. "Variational Methods for Image Segmentation", Birkhauser Verlag. 1995
11. H. Seo, N. Magnenat Thalmann."An Example-Based Approach to Human Body Manipulation", Graphical Models, Academic Press, Vol. 66, No. 1, pp. 1~23., January 2004
12. H. Seo, N. Magnenat-Thalmann, "An Automatic Modeling of Human Bodies from Sizing Parameters", ACM SIGGRAPH 2003 Symposium on Interactive 3D Graphics, pp19-26, pp234, 2003
13. L.Vacchetti, V.Lepetit, G.Papagiannakis, M.Ponder, P.Fua, N.Magnenat-Thalmann, D.Thalmann "Stable Real-Time Interaction Between Virtual Humans and Real Scenes", 3DIM, pp. 449-456, Banff, Alberta, Canada, 2003
14. Mun Wai Lee and Isaac Cohen**,** Human Upper Body Pose Estimation in Static Images**,** Institute for Robotics and Intelligent Systems**,** Integrated Media Systems Center, University of Southern California Los Angeles, CA 90089-0273, USA
15. I.A. Kakadiaris, D. Metaxas, 3D Human body model acquisition from multiple views, in: Proceedings of the Fifth International Conference on Computer Vision, pp. 618–623, Boston, MA, June 20–23,1995.
16. Raquel Urtasun and Pascal Fua, 3D Human Body Tracking Using Deterministic Temporal Motion Models, Computer Vision Laboratory, EPFL, CH-1015 Lausanne, Switzerland