ESPRIT LTR Project 9036 CHARM

Replacing Deliverable 11

Test sequences for validation

Generation of "real" images and 3D+ t reconstruction

Report

J Blat - V Caselles - J Fuster - J Verdera - G Bernat

Universitat de les Illes Balears

with the collaboration of Ll Muntaner (Hospital Son Dureta)

Test sequences for validation. Generation of "real" images and 3D+t reconstruction

Replacing CHARM D11 deliverable

First version. Report - Universitat de les Illes Balears

Contents

1. Introduction

The CHARM project fixed initially as one of the deliverables (to be provided by Hôpital Saint Louis, HSL) the generation of test sequences for validating the results of the project. After the failure of HSL in the project, this subtask was assigned to University of Geneva in collaboration with their medical consultants. First, work was done for getting MR images, as scheduled in the programme, and they were used for preliminary trials for 3D reconstruction of anatomical elements. But the approach to use these images for validation in the project was finally abandoned because the current available technology for MR images did not allow for 3D movements to be obtained in any reasonable way (approximately 30 minutes are currently necessary to obtain a 3D static picture, making both prohibitive and impossible to get a 3D sequence). In the general CHARM meeting of January 96 it was decided that UIB should try to obtain some type of images of movements to attempt a first validation of the CHARM results with real images as well as to develop the methods for the validation and test them.

UIB, jointly with Hospital Son Dureta consultants, designed experiences which could allow to obtain simultaneous fluoroscopy (X-ray sequences) and video images; the fluoroscopy images should allow for getting the bones and their movement, while the video sequences should allow to reconstruct the skin surface movements.

These images were to be taken from different points of view and, based on stereoscopy methods, after camera calibration they could result into 3D-reconstructed sequences, which could then be compared and validated with the synthetic animated images simulated with CHARM methods.

In this report we describe the setting of the experiences, the methods used for camera calibration and 3D reconstruction, both of video and fluoroscopy sequences, and the results obtained. This deliverable is complementary of D12, **Integrated module for analysis / synthesis / matching**, in which the computer environment developed for the different tasks is described, the validation environment is analysed, and the results of the comparison of the test sequences and the simulated ones are reported and discussed.

But first we discuss the validation questions as perceived in the context of the CHARM project and, specifically, in this deliverable.

2. CHARM validation issues

The aim of CHARM is to obtain better synthetic animation of humans by developing new methods, with the results being potentially useful to support medical practice (as well as for other applications, such as sports performance, training computer animation films, and entertainment). Different modalities or types of moving 3D images have to be analyzed and synthesized and compared in order to validate the methods of the project. More precisely, CHARM aims at developing a model of (a part of) a 3D solid human body to obtain more realistic movements and deformations coming from a physically-based simulation; the model is comprehensive as it includes not only geometric descriptions but topological structures, physical parameters, etc., and the results obtained by simulation are to be compared with real sources for validation.

The project has already developed a library of a part of the human body comprising models of the skeletal structure, made of geometrical composition, mechanical parameters for bones, mechanical definitions for skeletal joints, together with a hierarchical structure (the 3D geometric reconstruction of the surfaces has been based on the Visual Human Dataset and gives somehow a new type of anatomic atlas); and also models of the muscles and soft-tissues, which include topological modeling of the musculature, modeling of the muscle contraction force and of the constitutive equations of soft-tissues together with finite element modeling and procedures for both of them so that dynamical simulation of muscles activity and deformation is possible. Models of high-level of motion control have been developed too, in order to provide natural trajectories for automatic animation from a high-level specification, such as target configuration or end-effector positioning.

The shoulder-arm complex is currently the main focus of the research as it offers a quite substantial test-bed for the different ideas outlined above. The comprehensive model is meant to be used to produce simulated synthetic animations to be compared both with real images and medically originated images.

The main background of the project partners and their primary orientation are strongly computer graphics related, while on the other hand, recently techniques originated in graphics are quickly becoming significant in clinical practice, for instance in the use of tomographic images in concert with each other, or in the correlation of 3D tomographic diagnostic images obtained from different modalities such as PET or MR, or from the same modality in different times. Both medical and computer animation applications were thus inviting for conceiving and developing a validation environment based mainly in visualisation, with some tools included to support this environment. And the environment should be a 3D+time one.

The initial aim for the validation tasks was to base them on test (MRI) sequences obtained early in the project, which should provide - through image analysis - speed parameters to be fed into the CHARM models (both the finite-element simulations and the motion-control based ones) and compare the images obtained from these simulations to the test sequences.

As stated earlier, discarding MR images as source of test sequences was made much later than envisaged in the planning and the obtention of test sequences has come at a very late stage in the project. Also, the initial target of obtaining from simulation fully 3D realistic sequences has been abandoned because the finite element based methods developed mainly by the partner Istituto Superior Tecnico are also currently prohibitively time consuming (only results for two muscles could be reasonably expected in the time frame of the project) and thus, no rendering of a 3D shoulder-arm (which needs many more muscles) could be performed.

Consequently, the experiences for generating the test sequences were developed for obtaining both bones movements and skin surface movements, but the final test sequences are only for bones. The validation takes place by comparing a similar bones movement (abduction) both synthetically generated - with approximate movements - and reconstructed from fluoroscopy as the source of the real images.

Thus, a validation very limited compared with the initial expectations has been actually realised. But insight has been gained by developing some methods and results for this part of the project, and the validation attempt is also a first solution for this problem complementing other CHARM activities. This progress constitutes an initial step for further advances in research and offers both a view of limitations and an opening of ways for future work.

3. Setting of the experiences for obtaining the test sequences

3.1 Introduction

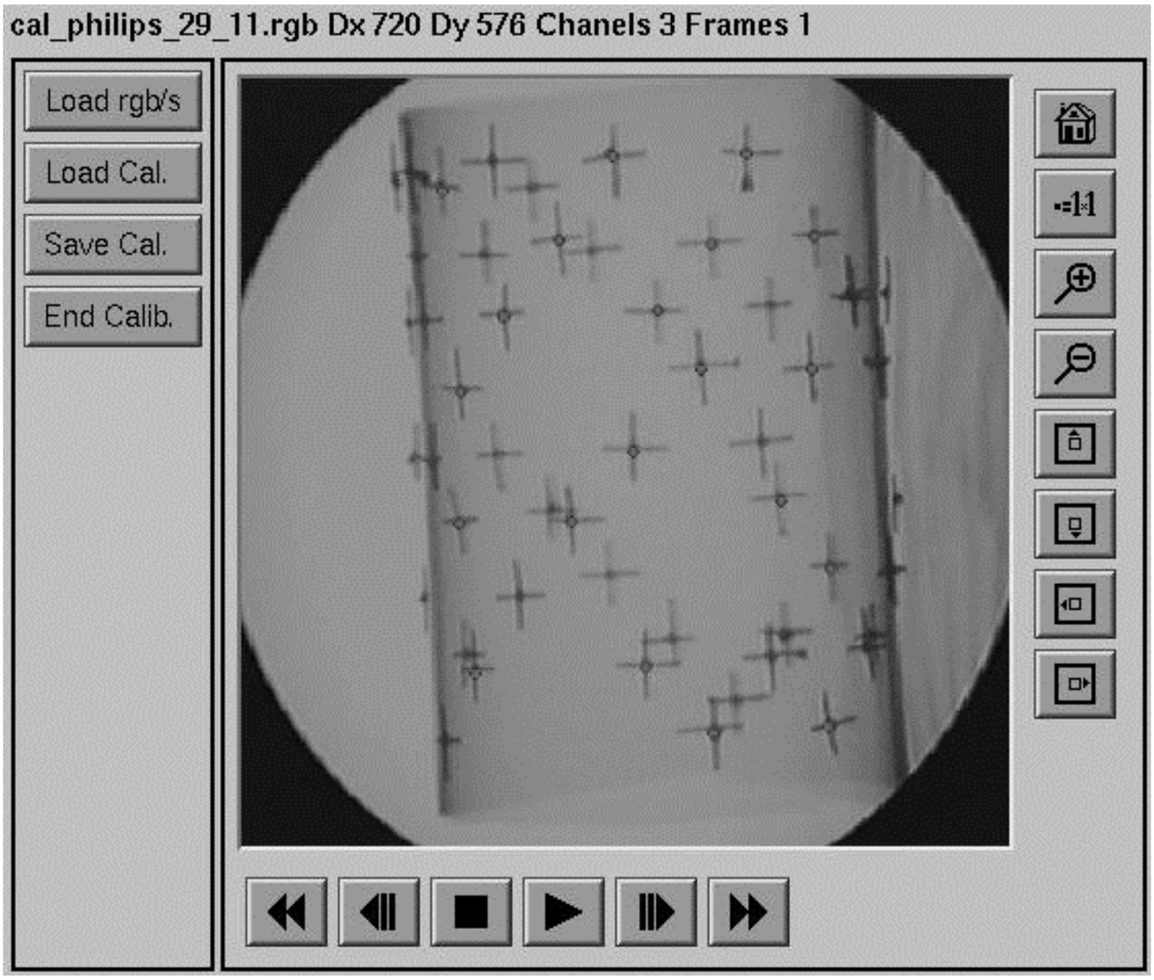
As stated above, from the test sequences we should get (from film analysis) kinematic parameters for running the simulations with the models (finite-element and control based) and then to compare these simulations with the testing ones. These sequences had to furnish, besides the speed parameters, 3D information of both bones and skin surface. The consultation UIB- Hospital Son Dureta decided then in favour of an experimental design for performing simultaneously fluoroscopies (X-ray sequences) and video sequences, the first ones for getting information for bone movement reconstruction and the second ones for the skin surface data.

To obtain 3D information from usual 2D images, different points of view are usually necessary, as it happens in our human stereoscopic vision system; two points of view is a minimum, although in order to obtain full 3D information (for instance, the front and the back) from video images of an opaque body, as is the shoulder-arm, more than two points of view are obviously needed. More points of view offer redundant information from which the precision of the 3D information can be improved and thus, in our experiences we used several video cameras for recording the movements and get this improved precision. As fluoroscopy is concerned, two approximately orthogonal viewpoints are enough, as there are no opacity problems and two X-ray devices were used for minimising irradiation.

The reconstruction methods that we have used require basically calibration, from which the camera parameters are obtained. After calibration, one puts into correspondence the same object point as seen in different images, which allows to compute the 3D position of the object point from the camera parameters.

For improving the video sequence processing, a sort of elastic wireframe was put on the skin surface, which should help to identify the corresponding points (this is a technique known as introduction of artificial markers). For the improvements which should come from redundant information, the markers should be fluoroscopy visible too.

The following image shows a fluoroscopy of the calibration object with the metalic markers visible:



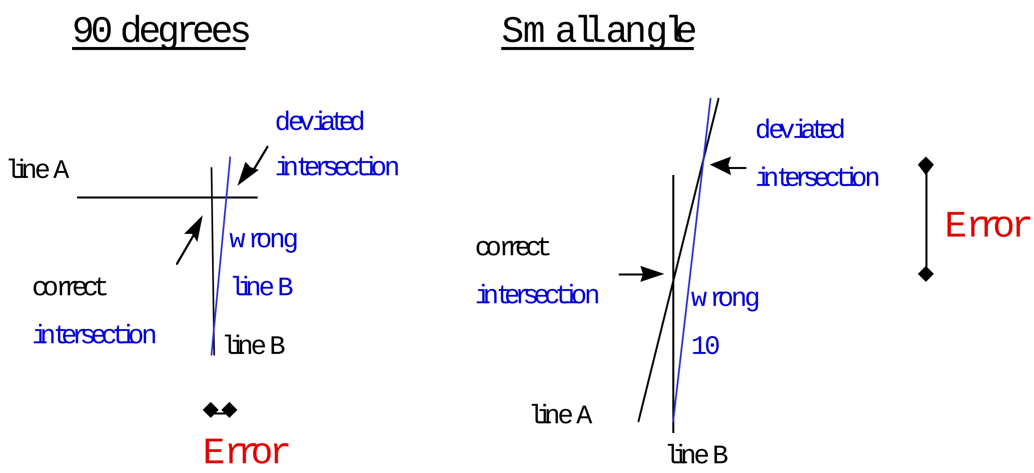
3.2 Preliminary trials with images (static and sequences)

The preliminary trials with virtual and real objects and cameras were extensive, as the result of these experiences should help to save efforts when the clinical setting experiences would be performed.

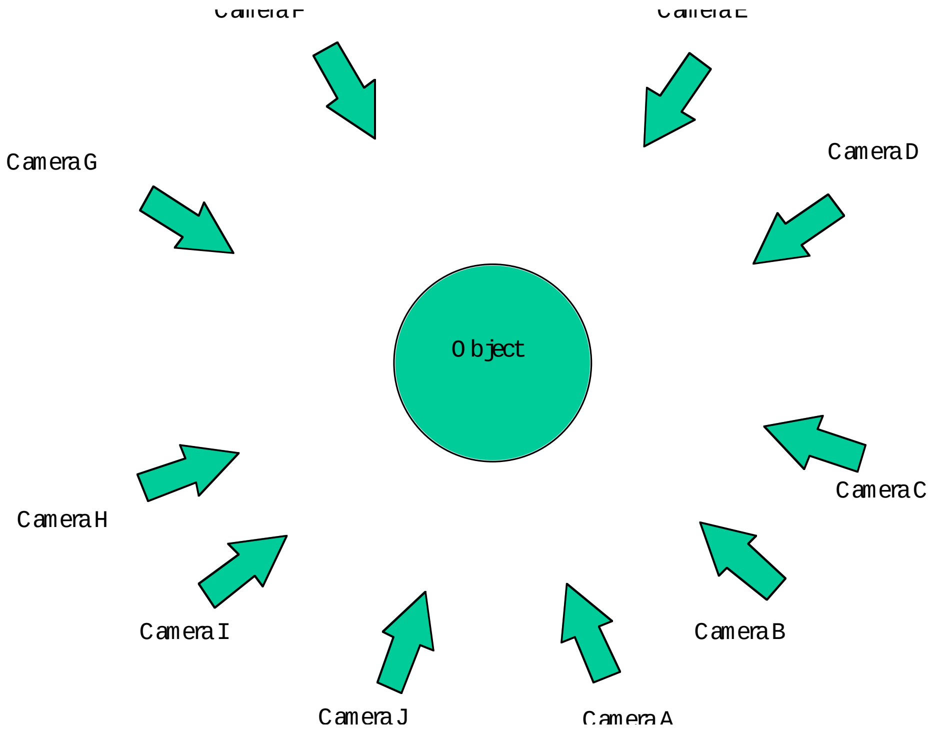
In order to test the algorithms and improve the methods, the first trials were performed with virtual objects and cameras; a synthetic cube and two virtual cameras were used to test two different calibration and reconstruction methods, both for checking the accuracy in the reconstruction and the efficiency of the methods, namely, their respective rate of convergence.

The second tests were performed with a cylinder covered with mm grid paper and different real cameras (with two, four, five cameras). The aims of these tests were threefold: to develop a calibration object and test it, to check which camera resolution was required for good reconstruction, and to see whether the relative positioning of the cameras had influence on the accuracy of the reconstruction. The SGI Indycameras were shown to be adequate, and thus, semi-professional Hi8 type video cameras were chosen for the experiments. Some ideas about the best angles for placing the cameras were obtained, and in the reconstruction process it was observed that automating matching of the corresponding points in different images was not very feasible. Some discussions about distortion are given later in the report.

The following diagram shows the error differences according to the orthogonality of the main axis of the cameras; orthogonality gives a big advantage:

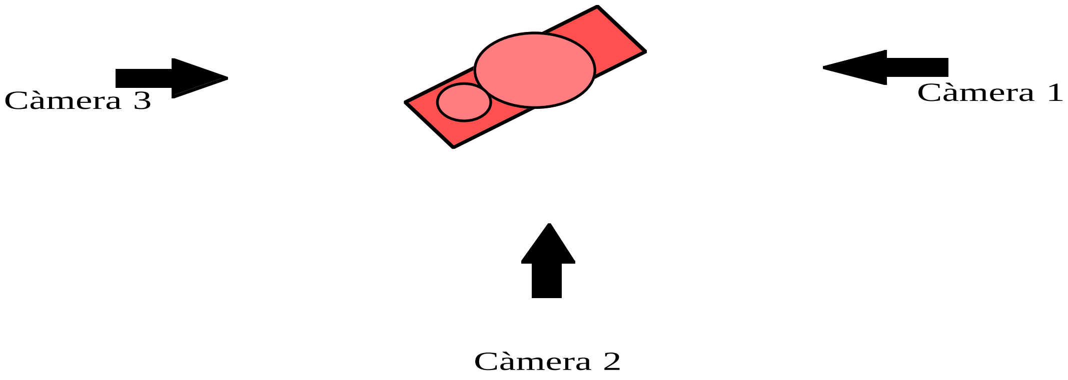


The following graphic shows one of the experiences with the calibration object in the middle and several cameras:



A third series of tests was performed for checking the robustness of the methods, by calibrating with one object and reconstructing another one, with several number of cameras. The final accuracy of the method recorded was about 5% of error in the estimates obtained, as compared with the actual grid measurements.

The following image shows the diagram of one of the experiences performed with two object and three cameras:

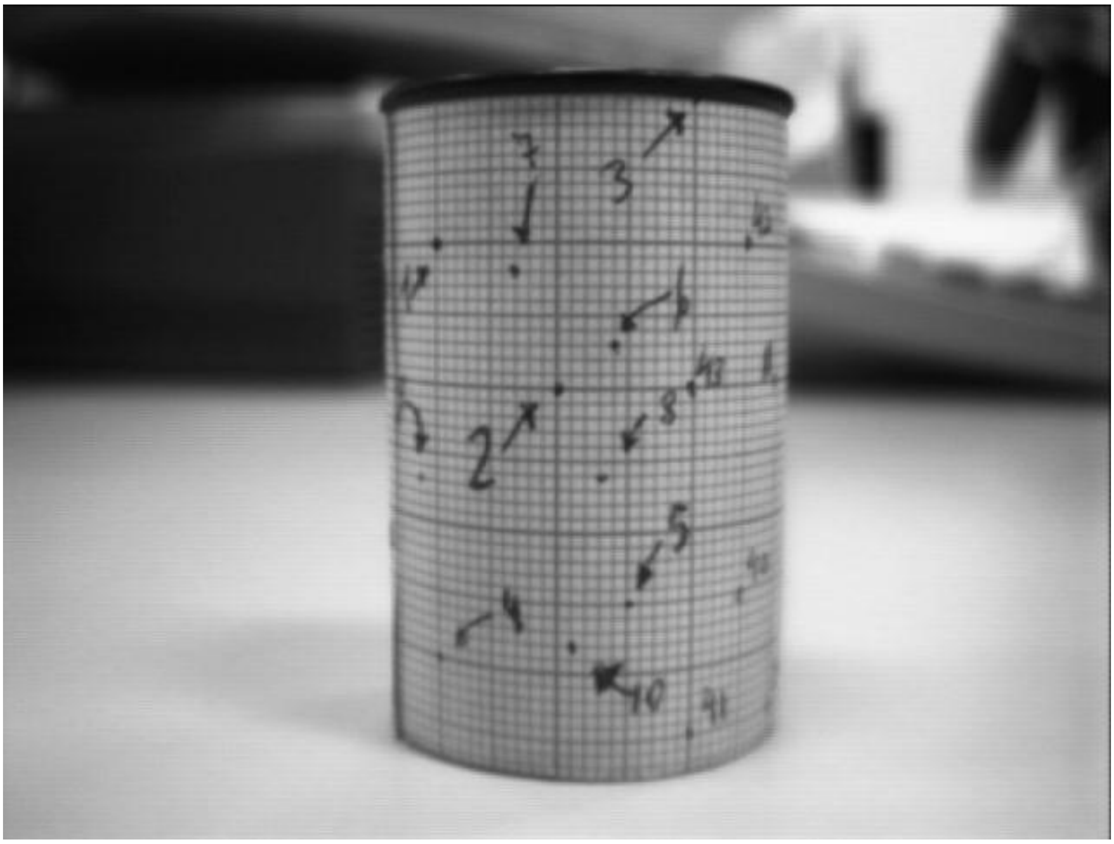


3.3 Medical experiences 1 and 2 (with and without video)

As mentioned earlier the clinical experience was to record simultaneously by fluoroscopies and video some shoulder-arm movements, with an elastic grid with markers on the skin surface in order to facilitate the process of putting in correspondence the conjugate points in the different images.

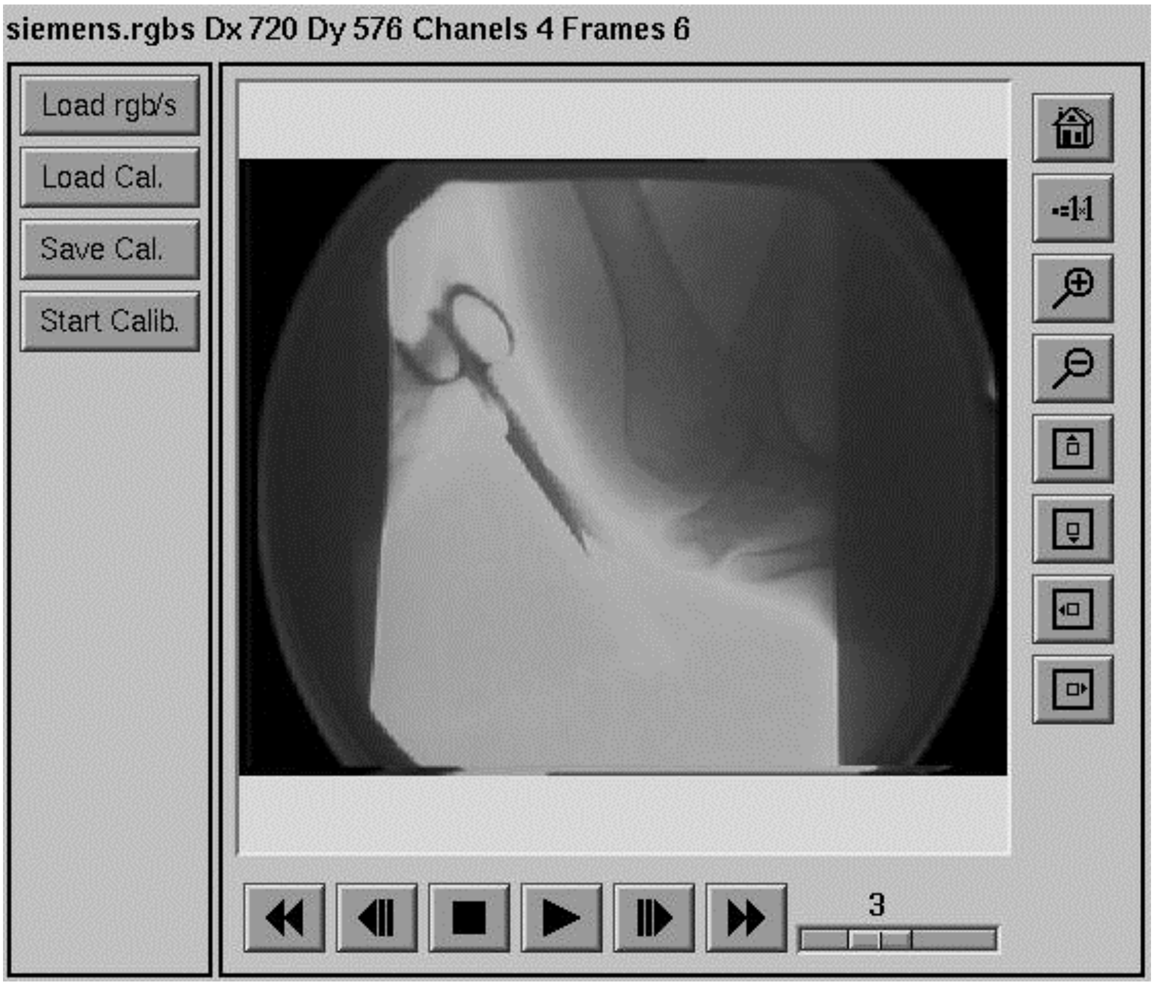
For the sake of the advantages of redundancy, the markers were to be visible both by ordinary video cameras and fluoroscopy. Initial tests were performed with the grid impregnated with the contrast solutions usually employed for getting visibility in fluoroscopies but with negative results because too high a radiation had to be used for the grid to become very slightly visible. Finally, the markers used were small metallic crosses sewn to the grid, which were visible both under fluoroscopy and ordinary video. The experimental design used for calibration a cylinder with a mm grid on the surface, where the same type of markers (i.e., metallic crosses) had been glued.

The following image shows the calibration object:



After several trials, the synchronisation was performed based in a pair of scissors performing the type of *claquette* movement used for separating camera shots in usual film shooting.

The following image shows a fluoroscopy of the synchronising event:



In an X-ray room of the Hospital Son Dureta, Palma, the first recordings in a clinical setting were made, after several trials. Two fluoroscopy cameras, one fixed and another movable (Siemens and Philips, respectively) were placed at approximately 90 degrees; this equipment was connected to standard VHS recording devices. Six standard video cameras, attempting to cover as much visible surface as possible as available due to the position of the subject of the experience, and to the space limitations of the X-ray room, were used. The setting of the experience is shown in the video deliverable of the whole project.

The fluoroscopy cameras, specifically the Philips one, could not cover the whole of the shoulder-arm, and an area composed of the shoulder articulation and approximately half of the humerus was selected. Standard procedures for information and consent of the voluntary subject were followed.

The first experience resulted in the recording of two movements, abduction and pronation-supination of about 5 seconds of duration each, and of the calibration object.

After analysis of the results of the first experience, which had quite a few problematic points, and in order to improve the images for the final validation involving only bone movement, a more specific and simpler second experience was designed and performed. Only the abduction movement and the calibration were recorded targeting at getting a better view of the shoulder articulation. This is the final test sequence which has been used for CHARM validation.

3.4 Digitisation

Digitisation of the final test sequences from the recording equipment was performed through SGI Indigo-video cards using JPEG compression in hardware (set at 80% quality), which allowed almost real time input of the images, while preserving a reasonable accuracy with respect to the objectives of the design.

4. Camera calibration and 3D reconstruction from stereoscopy

4.1 The camera model and the calibration of a camera

In this work we use the standard geometric model, which considers the images obtained from objects through perspective projection, of equations

$$x' = f\,x \,/\, z$$
$$y' = f\,y \,/\, z$$
$$z' = f$$

where it is considered that f is the focal distance from the principal point O to the image plane, (x,y,z) are the coordinates of the object point and (x',y',z') are the coordinates of the image point (using the standard convention that the Z-axis is the same as the optical axis from O orthogonal to the image plane. For convenience, homogeneous coordinates (x,y,z,1) are frequently used. (Most of the following developments follow closely the book Horaud R, Monga O: **Vision par ordinateur, outils fondamentaux**, Ed Hermès, Paris, 1993).

A transformation camera-image is used to convert from length (measured, for instance, in millimeters) to pixels, using as parameters the pixel coordinates of F, the projection of the principal point on the image plane (uo,vo,wo), the scale vertical factor (ku, pixels/mm) and the scale horizontal factor kv (also in pixels/mm).

Discarding the coordinate which is always fixed, we end up with the following final equation for this model:

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \mathbf{Ic} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \text{ where } \mathbf{Ic} = \begin{pmatrix} \alpha_u & 0 & u_o & 0 \\ 0 & \alpha_v & v_o & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

This model is thus defined by four parameters, called the intrinsic parameters: $\alpha_u = -k_u f$, $\alpha_v = k_v f$, u_o i v_o.

In order to recover 3D information, another transformation, the object-camera transformation is used:

$$\mathbf{A} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{R} & t \\ 0 & 1 \end{pmatrix}$$

Finally, the compact representation of these transformations from the object point (X,Y,Z) and the image point (u,v) is given by

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \text{ or } \mathrm{M} = \begin{pmatrix} m_1 & m_{14} \\ m_2 & m_{24} \\ m_3 & m_{34} \end{pmatrix}$$

with $m_i = (m_{i1} \quad m_{i2} \quad m_{i3})$. By identification, the intrinsic and extrinsic parameters can be computed through:

$$r_3 = m_3$$
$$u_o = m_1 * m_3$$
$$v_o = m_2 * m_3$$
$$\alpha_u = -\| m_1 \wedge m_3 \|$$
$$\alpha_v = \| m_2 \wedge m_3 \|$$
$$r_1 = 1/\alpha_u (m_1 - u_o m_3)$$
$$r_2 = 1/\alpha_v (m_2 - v_o m_3)$$
$$t_x = 1/\alpha_u (m_{14} - u_o m_{34})$$
$$t_y = 1/\alpha_v (m_{24} - v_o m_{34})$$
$$t_z = m_{34}$$

Calibrating the camera is the procedure which allows to compute these parameters from measures. There are twelve unknown camera parameters while each image point corresponds to an object point through the following two equations:

$$u = \frac{m_{11} * X + m_{12} * Y + m_{13} * Z + m_{14}}{m_{31} * X + m_{32} * Y + m_{33} * Z + m_{34}}$$

$$v = \frac{m_{21} * X + m_{22} * Y + m_{23} * Z + m_{24}}{m_{31} * X + m_{32} * Y + m_{33} * Z + m_{34}}$$

Thus, a minimum of six points is required to perform the calibration. The resulting system of equations is linear and homogeneous. Usually, the parameter m34 is fixed to one, and thus the parameters are computed up to a factor. The solution is also equivalent to replacing the zero right hand side by an error and getting the optimal solution by least squares. But we have used the Faugeras-Toscani method, where the module of m3 is equated to one (which is a natural geometric constraint). Taking again the criterion of error optimisation, and using constrained optimisation, the solution is equivalent to obtain the minimal eigenvalue of a matrix. Yet, there is another question, which is the sign which is equivalent to deciding whether the calibration object is in front or behind the camera; of course, the former decision is taken.

This approach is valid when redundant information is used to obtain estimates with less error, i.e., more than six points are used; actually, if n points are used, 2n equations are obtained only twelve unknowns and the overdetermination is again solved by least squares (using Faugeras-Toscani condition), and the minimal eigenvalue of the corresponding matrix is seeked.

In order to compute this minimal eigenvalue, the inverse power method, derived from the power method is used. If we look for the largest module eigenvalue of a matrix D, the iterations $x^{(k+1)} = Dx^{(k)}$ $(k \geq 0)$ (for any initial vector) verify the following property, the quotients $q_i^{(k)} = \frac{x_i^{(k+1)}}{x_i^{(k)}}$ $(i = 1,2,3)$ have as limit the "largest" eigenvalue λ_1.

For symmetric matrices, as it is this case, the Rayleigh quotients $\sigma_k = \frac{x^{(k)T} D x^{(k)}}{x^{(k)T} x^{(k)}}$ $(k \geq 0)$ are more efficient and have as limit the same eigenvalue.

In our case, as we are interested in the minimal eigenvalue, we compute the largest eigenvalue of the inverse matrix, D^{-1}; this is called the inverse power method.

4.2 3D reconstruction from images from different viewpoints

If we have two points (u,v) and (u',v') in two images (obtained with two calibrated cameras) corresponding to the same object point with coordinates (X,Y,Z), we have the following equations

$$u=\frac{m_{11}*X+m_{12}*Y+m_{13}*Z+m_{14}}{m_{31}*X+m_{32}*Y+m_{33}*Z+m_{34}} \quad v=\frac{m_{21}*X+m_{22}*Y+m_{23}*Z+m_{24}}{m_{31}*X+m_{32}*Y+m_{33}*Z+m_{34}}$$

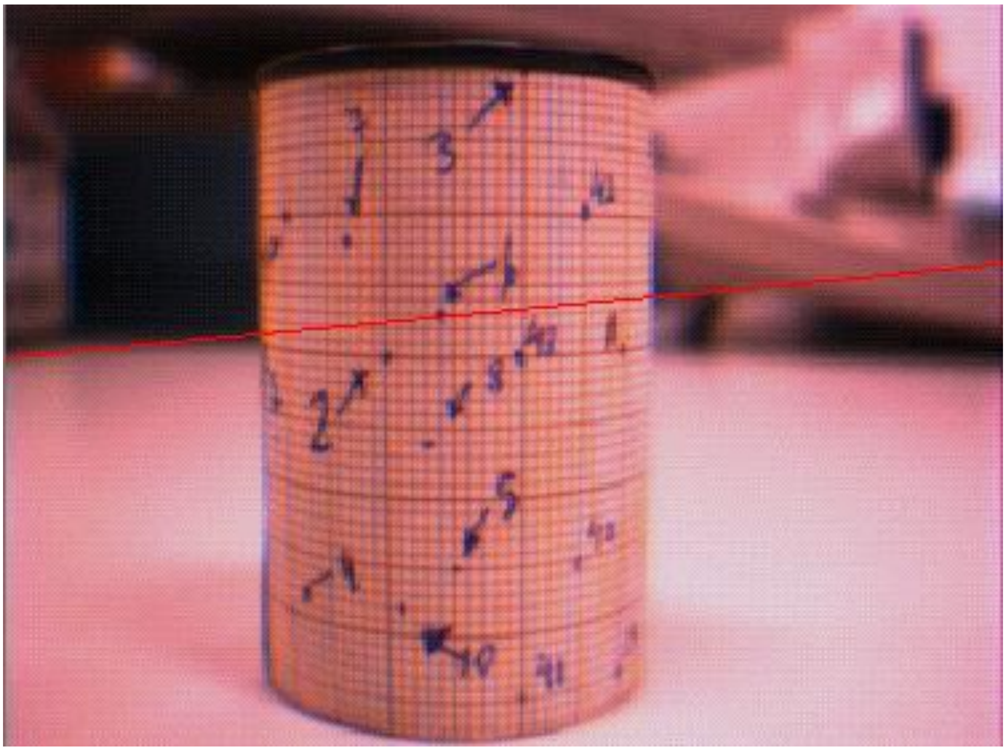
$$u'=\frac{m'_{11}*X+m'_{12}*Y+m'_{13}*Z+m'_{14}}{m'_{31}*X+m'_{32}*Y+m'_{33}*Z+m'_{34}} \quad v'=\frac{m'_{21}*X+m'_{22}*Y+m'_{23}*Z+m'_{24}}{m'_{31}*X+m'_{32}*Y+m'_{33}*Z+m'_{34}}$$

(or, in a more general case, if we have n images of the same point obtained with n calibrated cameras, 2n equations); now the unknowns are the three coordinates of the point. Again, using least squares and a standard Gauss method, the coordinates of the point are obtained.

4.3 Conjugate points and epipolar lines

It is well known that the conjugate point of a given one has lie in a straight line, called the epipolar line. Once a point is selected in an image, the epipolar line can help in the other image can help in the selection of the conjugate point. Thus, the possibility of drawing the epipolar lines was introduced in the environment.

The following picture shows an epipolar line drawn on the calibration object:



4.4 About distortion

The images from fluoroscopy do not have any detectable distortion. Methods for correcting the slight radial distortion appearing when using standard video cameras, were developed following Tsai R Y (A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses, **IEEE Journal Robotics Automation**, 3(4), 323-344, 1987) - radial distortion is the standard one to be considered - . An extra camera parameter has to be added for the calibration, but the resulting equations are non linear. An over-relaxation algorithm was used to solve these overdetermined equations. The results obtained after distortion correction were practically the same ones as without correction and in the final experiences no correction was introduced.

5. Calibration and reconstruction of the test sequences

5.1 Calibration

Although there are more advanced methods of calibration, it was decided to use an object with known measures (a cylinder with a mm grid paper glued on the surface and with metallic markers) in order to improve the accuracy and robustness of the method. During our experiences, no problems were perceived with this calibration method.

5.2 Reconstruction of skin surface through points

The correspondence problem of determining the conjugate points in different images is known to be a difficult one. Apart from operator driven methods, some automatic methods for grey level images are based on correlation of regions, others are based on the correspondence of grey levels and some use the edges of the objects.

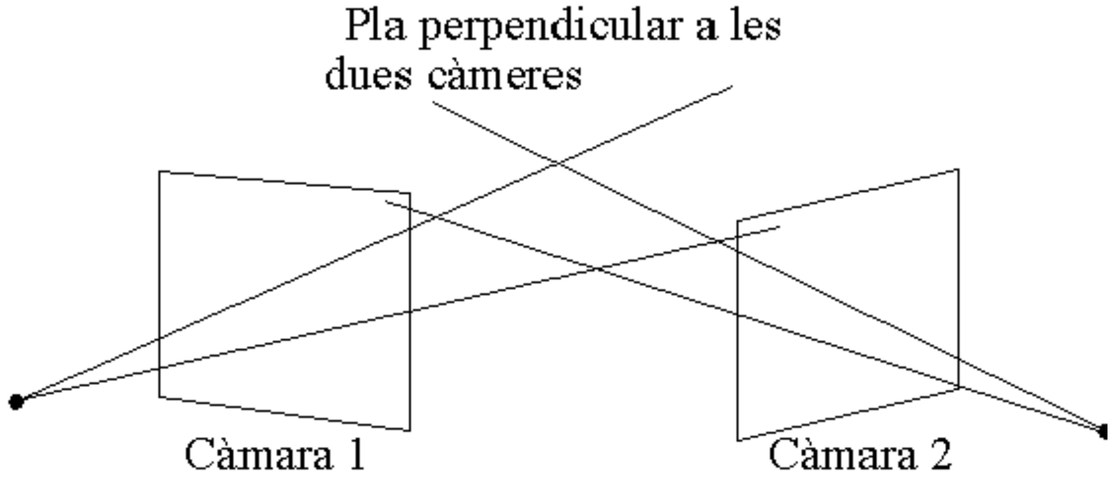
In our case we chose the precise identification of points (facilitated by the markers) by operators, which is very time consuming, but less prone to errors. The determination was done frame by frame in the sequence and no temporal related information was used. Even then, the determination of the conjugate points in the images is quite difficult. Thus, it was decided to use the epipolar lines in order to help the operator. From the experience, it was seen that following the path of the points in the scene helped the determination of the conjugate points.

The determination of the conjugate points and the use of the camera parameters obtained by calibration allowed the obtention of the spatial 3D coordinates of the corresponding object points. Transforming these points in surfaces is not a trivial problem: triangularisation requires the points to be ordered. Instead of developing our own methods, a public domain software (Quick Hull) for getting the convex hull of the points was used. This was not a very important limitation because for the validation basically the inertia axis of the surfaces were used, and these were computed with a reasonable degree of accuracy. Anyway, the skin surface reconstruction methods should require much more development if the same experimental design is to be used and skin surface movements have to be precisely determined.

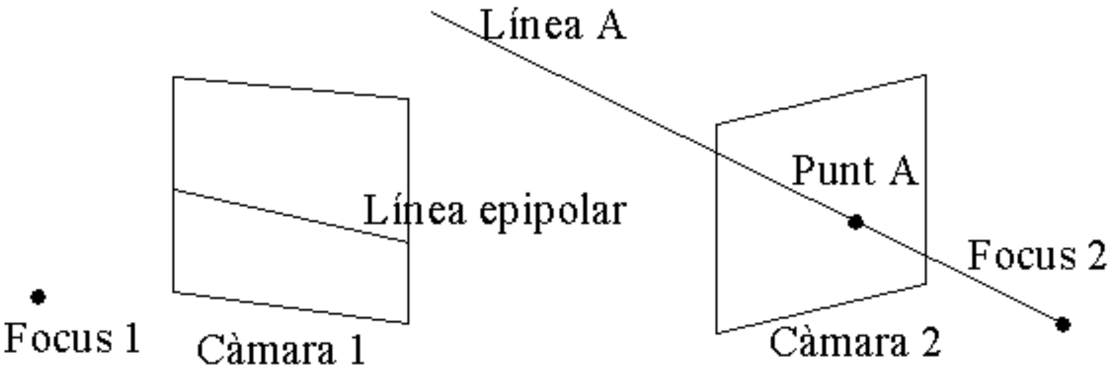
5.3 Bone reconstruction

In order to reconstruct the bones, the procedure aimed at getting approximate sections of the bones. This reconstruction is somehow more difficult because on the two fluroscopies there are no points marked previously on which to base precise point by point reconstruction. Thus the approach has to be changed completely in order to perform bone reconstruction.

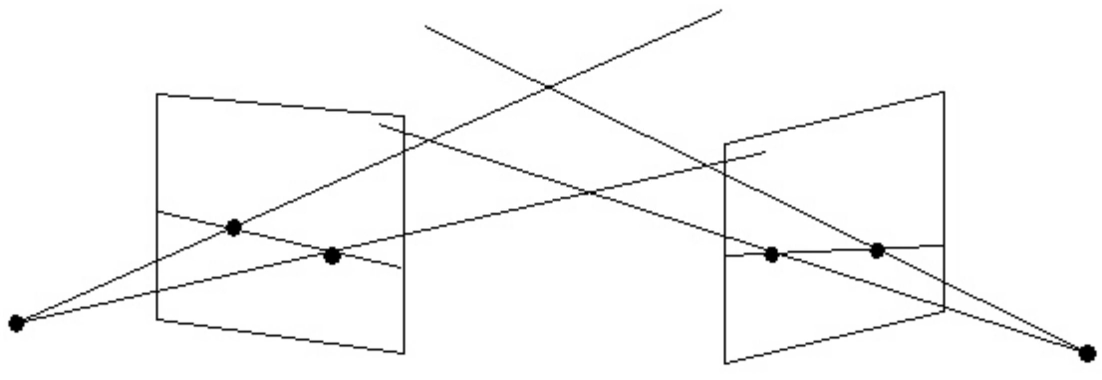
Thanks to the property of images corresponding to points of view which are almost orthogonal, we can determine a plane orthogonal to both cameras, which makes up a section of the bone.



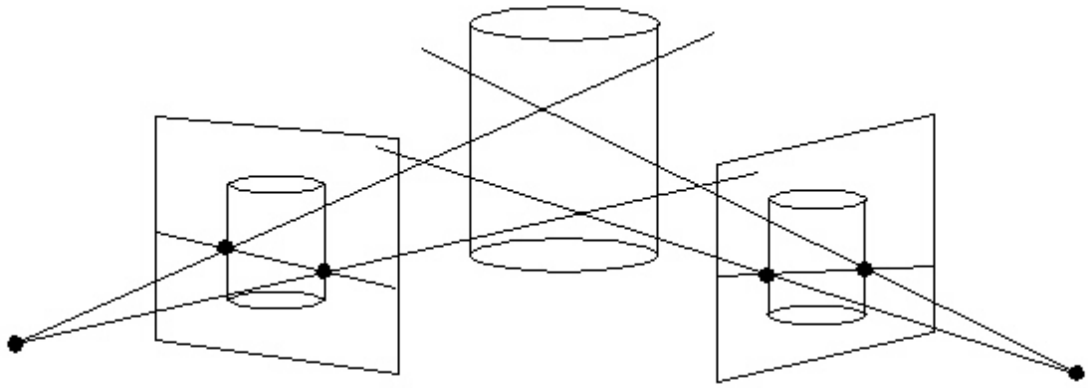
This plane is easily obtainable through the epipolar lines: where conjugate points lie, represented in one camera by the line determined by one point from the other image and the focus of the other camera. See the following diagram



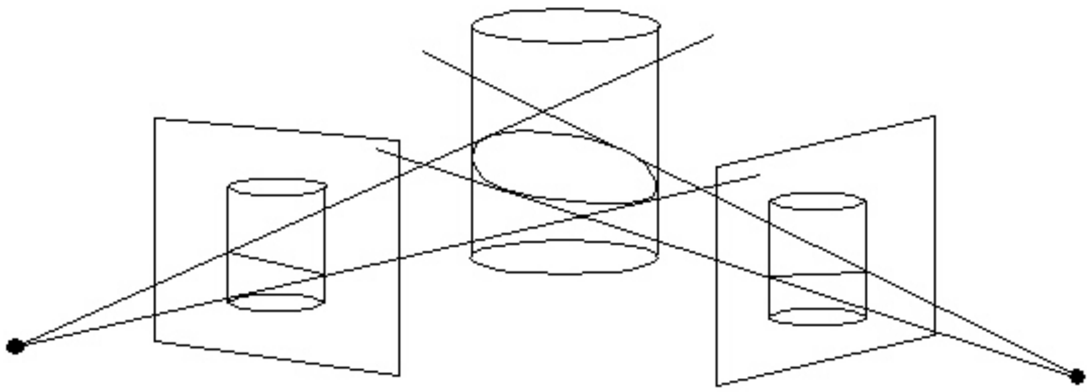
Then, in order to obtain an orthogonal plane to the two cameras we only need one point of an image in one camera which determines the epipolar line on the other camera and any point in the epipolar line defines an orthogonal plane to the two cameras, as seen in the following diagram:



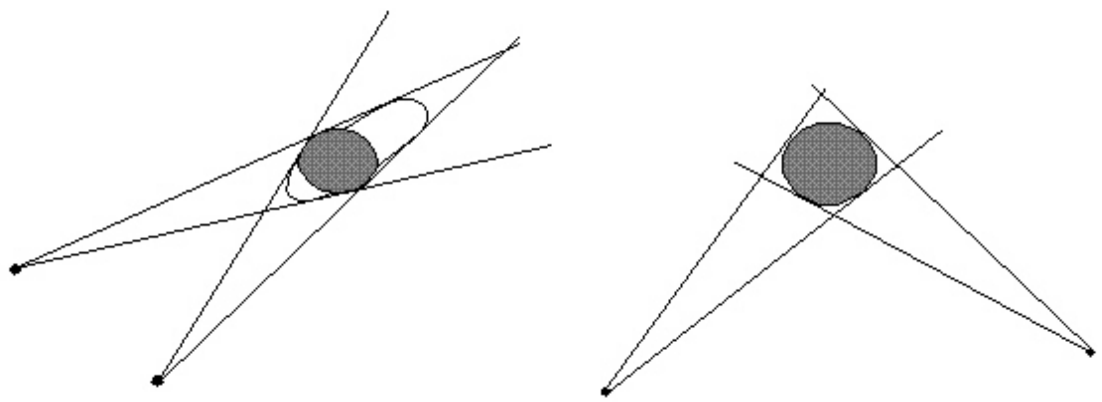
Then a parallelepiped containing the object to be determined is easily drawn,



and the curves corresponding to the object can be obtained



Thus we approximate the bones through their sections, based on the special fact that fluoroscopies are practically orthogonal. If the points of view are not orthogonal quite large errors can be incurred into, as shown in the following diagram



6. Conclusions and perspectives

6.1 The experimental design

The use of fluoroscopy and video images was decided as an emergency solution to try to develop the validation methods for CHARM and at a quite late stage of the project trajectory. On the other hand, developing this experimental environment has proved to be, as it usually is, quite complex in time, resources, tests - novelty in experiments always requires them. This is in itself difficult and made more difficult because the use of a reasonable amount of medical resources had not been scheduled in the project. Thus, the results are somehow *ad hoc*, and the resources devoted and required for the stabilisation of this environment are probably not justified unless more direct clinical applications are envisaged.

Although a direct application for the study of recidival shoulder problems was envisaged from the development of the experiences, only resources devoted to consolidation of a more stable environment could lead to this clinical applicable study.

Another alternative to this experimental environment is likely to be the use of new MRI techniques which are becoming commercially available. This would lead to a less invasive technique; and the majority of CHARM methods - as used in the reconstruction based on the VHD - will be more suited to this environment.

6.2 The reconstruction methods

The fluoroscopy images have shown to be much more useful than the video images for the reconstruction methods that were designed for the skin surface. Even with markers, the determination of the conjugate points was quite complex. A much deeper development of the correspondence problem (including the use of temporal information) would be required for significant improvement.

The reconstruction performed on the bones has been easier to develop due to the better images and the more targeted approach. The procedure through which sections have been reconstructed with the help of epipolar lines is prone to automatisation with little effort as it basically involves the intersection of the bone edges with the epipolar lines.

Last but not least, the little time available has not allowed to explore the use of temporal information. The methods are basically static, considering only isolated frame information. It is clear that significant improvements should be achievable by considering the sequences.

6.3 The test sequences

Due to *ad hoc* character of the environmental setting developed, and the limitations underlined above, longer time would be required to make it more stable and hence more useful. From the different trials and clinical experiences, only the last one could be described as really useful for testing and for more use than the actual final CHARM validation.

The other sequences still gather the *ad hoc* character described above, are essentially trials, and cannot be considered as suitable for external use, more than what has been described in this report.