Real time positioning system based on active stereo and facial features identification

Iuri Frosio[1,2], Raffaella Lanzarotti[1], Paola Campadelli[1], Giancarlo Ferrigno[2], N. Alberto Borghese[1]

1 Department of Computer Science, University of Milan, Italy [frosio, lanzarotti, campadelli, borghese]@dsi.unimi.it

2 Bioengineering Department, Politecnico of Milan, Italy

Abstract

We propose a new system for patient's head positioning for radiology and radiotherapy. The system is constituted of two low-cost digital cameras and a digital projector. As a first step a bright pattern is projected onto the subject's face and a 3D model of the face is reconstructed in real-time. The pattern guides the selection of feature regions of the face: the eyes, the lips and the nose. Inside these regions features are recognized using image processing techniques; 3D many-fold analysis can improve this stage. From these features the position and orientation over time of the face can be determined.

1. Introduction

High accuracy positioning of a patient with respect to the instrumentation is fundamental in many clinical applications, in particular in dental radiology and radiotherapy[4,5]. Traditional radiotherapy positioning systems make use of a mask to lock the patient's head position before irradiating it[1]: they allow to reach high accuracy, but they are extremely uncomfortable to the patient. In dental radiology, light beams are projected on patient's face in order to facilitate a "hand made" positioning procedure: the clinician has to make the lines passing over body reference points. This method is neither fast nor accurate.

For these reasons, very recently, image processing and computer vision techniques are being explored to make patient's positioning, an automatic procedure: fast, accurate, reliable and without lack of comfort for the patient[2,6]. Video images and subtracting techniques are profitably used in[6], obtaining residual error of 1-3 mm. In[2] a positioning system which uses stereo images, laser lines and markers on patient's head is presented; residual error within 1 mm are reported. However, both methods require the acquisition of reference images or the reconstruction of a 3D reference surface: position is computed comparing the reference surface with the actual surface, which cannot be performed in a short time.

We report here on a pilot project, aimed to equip panoramic radiographic systems (e.g. Dentsply Orthoralix™, Siemens Orthopos™) with a low cost automatic positioning system, which will substitute "hand-made" positioning. Aim of this system is to determine the relative position and orientation of the face with respect to an external reference system, starting from nine feature points (eyes corners, mouth corners and middle points, nose tip), automatically identified on a pair of images of the patient's face. Using features instead of reference surfaces or images, has allowed us to develop an algorithm, which does not require iterations and therefore is suitable to real-time implementation.

2. Material and methods

1 – System overview

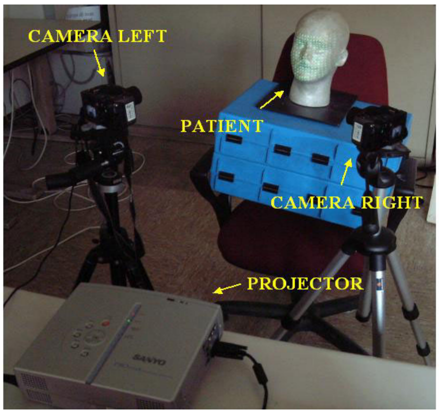

Figure 1 *– System setup.*

The system is constituted of two commercial digital cameras (1 x Fujifilm FinePix S602Zoom, 1x Fujifilm

FinePix 4900Zoom) and one digital slide projector. The latter is positioned in front of the patient, with the two cameras on its sides forming an angle of about 40° (Figure 1) and it projects a luminous pattern, constituted of a matrix of spots, over the subject's head.

First step is the construction in real-time of a 3D digital model of the subject's head (Figure 3). Afterwards, from the 3D model and the two images a set of feature points are automatically identified and their 3D position are used to estimate the orientation and location of the subject's head.

2 – Calibration

Before using the system, calibration has to be carried out. This is the procedure, which estimates the geometrical and optical parameters of the surveying system. The photo cameras are calibrated by taking the image of a chessboard surveyed in different orientations and positions: the internal, external and distortion parameters are estimated for each camera [3]. The projector does not need an explicit calibration as the direction of the light beams is the only information required. Therefore, we have developed an innovative calibration procedure, termed implicit calibration, in which the position and orientation of the straight lines associated to each light beam, are explicitly computed. To the scope, a planar white sheet is positioned at different locations and attitudes, in front of the system, and surveyed by the pair of photocameras. The spots projected over the sheet by the slide projector can be easily identified on the pair of images, and their 3D position can be reconstructed. Each light beam can be identified by fitting a straight lines to the 3D positions of the spots produced over different sheets. This method has the advantage not to be sensitive to distortions produced by the slide projector.

3 – Recognition of the beam spots

Once the patient is in position, first two pictures of the patient's head (one for each camera) are taken. Afterwards, the projector is switched on to take other two other pictures (Figure 3). The spots on patient's face can be identified by subtracting the images with and without the light pattern. A slight underexposure of the cameras makes the process more reliable.

4 – Points matching & Model construction

To reconstruct the 3D points, the spots have to be first matched between the images pair; a robust procedure has been developed to the scope.

For each luminous spot D_L on the left image, the corresponding epipolar line, e_R, is drawn on the right image; the corresponding spot D_R in the right image lies on e_R (and viceversa). Moreover, projecting each luminous beam $B_{i,j}$ on the two images, a second pair of epipolar lines ($b_{Li,j}$, $b_{Ri,j}$) is obtained. Therefore the point associated to D_L (D_R) must belong to $b_{Ri,j}$; similarly, the point associated t D_R (D_L) must belong to both $b_{Li,j}$ and e_L (Figure 2). This represents a trifocal constraint and makes matching robust.

Nevertheless a significant number of false matching (10% - 40%) does occur due to the density of spots; trifocal constraint alone is not sufficient to find reliable matches. To make matching more robust, we have added a neighborhood constraint, developing a novel algorithm, which we have termed "Four-links Matching Algorithm". The basic idea is to match groups of five ordered points, visible on the pair of images (Figure 3).

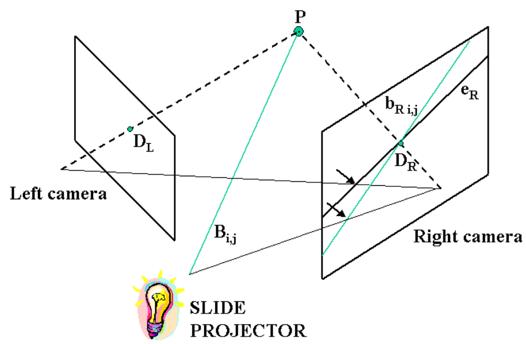

Figure 2 *– Multipolar geometry of the active stereo system – the two arrows indicate the two epipoles of the right image: the upper one is referred to the perspective geometry of the two cameras, the lower one to that of the right camera and the projector.*

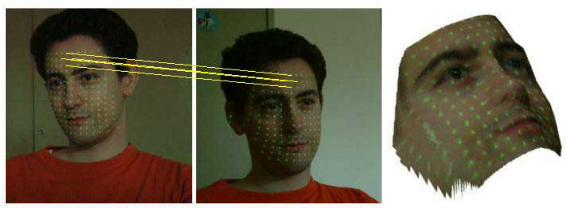

Figure 3 *– Four Links Matching and the colored reconstructed surface (only points that satisfy the trifocal constraint are shown).*

First of all, each spot D_L (D_R) on each image is assigned to one or more luminous beam $B_{i,j}$, on the basis of its distance from $b_{Li,j}$ ($b_{Ri,j}$). A connectivity matrix is built for all the spots on each image as follows. 2D Delaunay triangulation is performed and the resulting mesh, connecting the spots, is analyzed. Two spots, which constitute the vertices of the same triangle, are assumed to be a *link*. When, for a single spot, s_i, four links are found, matching for the entire set (s_i plus the four linked spots) is evaluated (*"four-links matching"*). To the scope, the set is matched with the other four-links structures found on the other image. This is made possible by the set-up used; this guarantees that links orientation is similar on the two images, and it is similar also to that of the segments joining two adjacent straight line beams from the projector. Delauney triangulation allows to make matching more robust than simply considering neighbor spots as it allows discarding degenerate configurations arising from spots not-surveyed because of occlusions.

Occlusions may prevent both cameras from seeing all the spots (e.g. the spots on the cheeks or on the lateral aspect of the nose). However, the 3D position of these spots can still be reconstructed by intersecting the straight line associated to the corresponding beam with that through the spot image and the projection center of the camera.

Once the 3D position of all the spots has been reconstructed, the 3D model of the face can be built by filtering the data through a regularizing surface[7] or, more simply, connecting the ordered points. As beam spots are quite accurate (reconstruction errors are within 1mm in 3D space), mesh filtering is not required. Finally, the two images can be mapped over the surface parameterizing them through the calibration parameters[11].

5 – Features location

As the patient stays always approximatively in the same position with respect to the instrumentation, a rough approximation of the features (eyes, mouth and nose) over the mesh and in the two images is available. This allows searching locally for the features, by using sophisticated image processing algorithms on small sub-images.

In particular, eyes corners are identified through a two step procedure[8]. First, the Hough transform is applied to the luminance horizontal gradient of the eyes image, and the pupil contour is identified by fitting a circle to the points over-threshold. Then, the luminance vertical gradient is analyzed and a deformable template[14] is fitted to the gradient associated to points inside the height interval determined by the pupil. The two joining points of the upper and lower template represent the eye corners.

To determine the lips border, clustering is first applied to the luminance vertical gradient of the mouth sub-image, in order to isolate the dark region corresponding to the lips cut. After having removed these pixel, a second clustering is applied to the CIE-LUV colored image to isolate the lips from the surrounding skin. A second order curve is then fitted to the boundary pixels for the upper and lower lip, and the intersection point of the two curves assumed as mouth corner.

Nose is characterized on the images, by two dark regions associated to the nostrils and a bright spot, associated to nose tip. Nostrils and nose tip can therefore be identified by using clustering on the nose sub-image. The middle point between the nostrils gives a good estimate of the horizontal position of the nose tip, while the vertical position is identified as the lowest position of the bright spot (Figure 4).

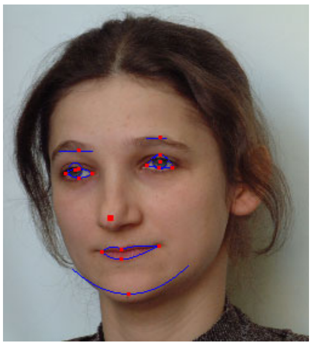

Figure 4 *- The features identified through image processing: the nose tip, the middle point and the corners of the mouth, the eyes are used as features.*

With the same procedure, also the middle point of the eyebrows and of the chin can be identified. However, these features are identified less reliably and will not be used for positioning purposes.

Once the points have been identified, the patient's head position is estimated through standard algorithms [9]. This can be input to the motorized radiographic machine for automatically correcting its position and orientation.

3. Discussion and Conclusion

The 3D model is not required to be extremely accurate as it is used mainly to determine the local differential properties of the geometry, to help in finding the features. Therefore we have chosen to project points, which can be detected with high accuracy and processed in a few time. This approach is different from traditional scanners, which project lines or grids, to maximize the number of range points, but require also some computational time to produce the 3D mesh.

Surface reconstruction is performed with high reliability: four links matching allows to generate an ordered point-cloud, without outliers. Processing time to reconstruct a pattern of 300 spots (including surface smooth interpolation) is about 15 seconds for a pair of 1280x960 images, on a Mobile AMD Athlon 2400+, in Matlab 6.5 but the code can be highly optimized.

All the features are identified within an error experimentally determined of a few pixels. We are working in improving this figures by introducing the epipolar constraint and differential geometrical information derived from the local curvature of the 3D model[12]. Features detection requires on its own a few seconds in IDL interpreted language.

Overall the system has the potentiality to become a real-time robust positioning system.

References

1. C.P. Karger, O. Jakel, J. Debus, S. Kuhn, G.H. Hartmann, Three-dimensional accuracy and interfractional reproducibility of patient fixation and positioning using a stereotactic head mask system, *International Journal of Radiation Oncology, Biology, Physics*, Volume: 49, Issue: 5, April 1, 2001, pp. 1493-1504

2. Borg J., Ramaseshan R., Heaton R., Alasti II., Islam M., Application of 3d digital surface imaging for patient positioning, *International Journal of Radiation Oncology, Biology, Physics*, Volume: 54, Issue: 2, Supplement, October 1, 2002, pp. 316-317.

3. Z. Zhang., A flexible new technique for camera calibration, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 22(11):1330-1334, 2000.

4. Xie, Q., Dinamo, A., Soikkonen, K., Wolf, J., Mattila, K., Gong, M., Effect of head positionong in panoramic radiography on vertical measurements: an in vitro study, *Dentomaxxillofacial Radiology* Volume: 25, Issue: 2, April, 1996, pp. 61-66.

5. Grush, Jr., William, H., Steffen Gary A., Heading in the right direction? An innovative approach toward proper patient head positioning, *Medical Dosimetry* Volume: 27, Issue: 4, Winter, 2002, pp. 261-264.

6. Johnson, L.S.; Milliken, B.D.; Hadley, S.W.; Pelizzari, C.A.; Haraf, D.J.; Chen, G.T.Y., Initial clinical experience with a video-based patient positioning system, *International Journal of Radiation Oncology, Biology, Physics*, Volume: 45, Issue: 1, August 1, 1999, pp.205-213.

7. Borghese N.A. and Ferrari S. (2000), A Portable Modular System for Automatic Acquisition of 3D Objects, *IEEE Trans. Instrumentation & Mesurement*, Vol. 49(5), pp. 1128-1136.

8. R. Lanzarotti, N.A. Borghese, P. Campadelli (2002), Robust identification of fiducial points for the reconstruction of 3D Human faces from raw video sequences, *Proc. First IEEE International Symposium on 3D Data Processing and Transmission*, Padova, June 2002, pp. 690-693.

9. Faugeras, (1993), Three-dimensional computer vision, a geometric viewpoint, MIT Press, 1993.

10. Tommasini T. Fusiello A. Trucco E., Roberto V., (1998), Making Good Features Track Better, *CVPR 1998*, Santa Barbara, CA, USA, pp. 178-183.

11. Weinhaus F. and Devarajan V., (1997), Texture mapping 3D models of real-world scenes. *ACM Computing Surveys*, 29(4):325-364.

12. Proesmans M. and Van Gool L., (1998), One-shot 3D shape and texture acquisition of facial data. *Proceedins ICCV, 1998*, pp. 18-25.

13. Grün A., (1985), *Adaptive least squares correlation: a powerful image matching technique*, South African Journal of Photogrammetry, Remote Sensing and Cartography 14(3), pp. 175-187.

14. Yulle A., Hallinan P., and Cohen D., (1992), Feature extraction from faces using deformable templates. International Journal of Computer Vision, 8(2):99-111.