Fusion of video and motion data: engineering tasks and clinical applicability

Ferenc Pongrácz
NKTH, Budapest
pongraczf@t-online.hu

Abstract

Status of a new project – focusing on technical aspects of fusion of live video and data from optical motion tracking – is summarized. Also, the significance of development in two areas is marked out: (1) in clinical field of endoscopic surgeries as demonstrating and training tool; (2) in biomechanical research for accurate detection of human motion curves.

Keywords: video tracking, motion data fusion, endoscope, movement analysis

Introduction

The idea of representing video image and captured motion data in a common reference frame is not new[1,2,3,6] but, because of many application areas and latest achievements in technology, it's worth to give attention to related technical and clinical aspects.

As an important application, this article gives an introduction of tools for training and demonstrations in endoscopic surgery. These goals can be reached by integrating endoscopic video into surgical navigation environment and performing motion tracking of surgical devices and endoscope. In future extension, for internet-based training and demonstrations, a special network transfer can be developed for transmitting live video with surgical planning and motion data. The education of surgical skills is a common problem in endoscopic surgery[7] that supports the proposed project plan[8]. Personal communications with head physicians in local hospitals confirm that, inexperienced intervention frequently rearranges the anatomical structures during ENT (Ear Nose and Throat) surgeries that can cause large obstacles in acute cases. In case of surgical planning with imaging data (CT, MR), the client observers follow in real-time the trajectory of endoscope and identify the critical anatomical parts.

Another important area of fused (conventional) video and motion data is closely related to biomechanical research. Reconstruction of 3D human movement is difficult if predefined graphical models are supposed to match bone movements during gait analysis or study of any pathologic motion[4,5,9]. Correction of geometrical axes or use of constraints in modeling after experiments makes the interpretation of motion diagrams almost impossible[12]. Introduction of 3D registered video and fused motion data from bone sensors helps a lot in testing kinematical models. The misalignment of graphical models can be visualized and compared to real anatomy on video screen using different camera positions. On screen editing is possible in live and playback modes. After involving and registering diagnostic volume (CT, MR volumetric models or planar X-Rays) to tracker's space, the accuracy of kinematical models – during video screen editing – can be easily tested even on static diagnostic data.

Methods

Video calibration

Positioning of video frame into 3D navigation environment is performed by registration of camera space to the space of motion tracking device. This is possible between 3D orthogonal coordinate spaces; therefore it has to be preceded by conversion of visible (usually distorted) image plane into an undistorted image plane. This step is named as calibration procedure which means calculation of transform between the distorted and undistorted image planes. Distortion parameters are given by[11]:

$$j' = a_0 + a_1 u + a_2 v + a_3 u^2 + a_4 uv + a_5 v^2 \qquad (1)$$

$$k' = b_0 + b_1 u + b_2 v + b_3 u^2 + b_4 uv + b_5 v^2 \qquad (2)$$

where u, v are the coordinates on undistorted (ideal) image plane and) are the calculated values. Undistorted image plane is a calibration grid containing large number of markers. Error function between observed (j', k') and calculated coordinate

$$E = (j - j')^2 + (k - k')^2 \qquad (3)$$

in vector form:

$$E = \|j - Pa\|^2 + \|k - Pb\|^2 \qquad (4)$$

where j, k are the vectors which are from distorted image plane. P represents the matrix which is set from values of ideal undistorted image plane (calibration grid):

$$P = \begin{bmatrix} 1 & u_1 & v_1 & u_1^2 & u_1 v_1 & v_1^2 \\ 1 & u_2 & v_2 & u_2^2 & u_2 v_2 & v_2^2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & u_{m-1} & v_{m-1} & u_{m-1}^2 & u_{m-1} v_{m-1} & v_{m-1}^2 \\ 1 & u_m & v_m & u_m^2 & u_m v_m & v_m^2 \end{bmatrix} \qquad (5)$$

Distortion parameters if both coordinates are involved:

$$a = [a_0 \quad a_1 \quad a_2 \quad a_3 \quad a_4 \quad a_5] \qquad (6)$$

$$b = [b_0 \quad b_1 \quad b_2 \quad b_3 \quad b_4 \quad b_5] \qquad (7)$$

In case of ideal radial distortion (for many endoscopes) distortion equations (1, 2) can be modified:

$$r' = c_0 + c_1 u + c_2 u^2 + c_3 u^3 \qquad (8)$$

where u is the radial distance on undistorted plane and r' is the estimated distance after distortion. In this case the error between the observed (r) and calculated (r') values:

$$E = \|r - Pc\|^2 \qquad (9)$$

where P contains the ideal image coordinates:

$$P = \begin{bmatrix} 1 & u_1 & u_1^2 & u_1^3 \\ 1 & u_2 & u_2^2 & u_2^3 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & u_{m-1} & u_{m-1}^2 & u_{m-1}^3 \\ 1 & u_m & u_m^2 & u_m^3 \end{bmatrix} \qquad (10)$$

$$c = [c_0 \quad c_1 \quad c_2 \quad c_3] \qquad (11)$$

(4) and (9) can be optimized by statistical methods. Vectors (6, 7) and (11) can be calculated by SVD[11] (singular value decomposition) which accurately estimates video distortions for both conventional and endoscopic cameras:

$$a = (P^T P)^{-1} P^T j \qquad (12)$$

$$b = (P^T P)^{-1} P^T k \qquad (13)$$

$$c = (P^T P)^{-1} P^T r \qquad (14)$$

Target positions marked out on video image should be also transformed back to the tracker's reference space or diagnostic volume. It's possible if the inverse of distorting equations (1)(2)(8) are known. This calculation is difficult from the results of first, "distorter" step and only numerically possible. Therefore another approach is proposed. According to it, the inverted parameters are calculated in the same way as the forward values but with different assignments. The input coordinates or distances

(u, v), used in the first step, are replaced by values located on the distorted image (j, k, r). Afterwards, the parameters P depend on the distorted coordinates and the error functions are used to approximate the ideal, undistorted values ("undistorter" step). Finally, the optimization can be the same as it was during the forward calculation.

In practical implementation the endoscopes with oblique lens axis should be also calibrated *(Figure 1)*. Best approach is to use ready-made, special calibration modules (clamps). These contain the calibration grid and are able to hold endoscopes in a predefined position.
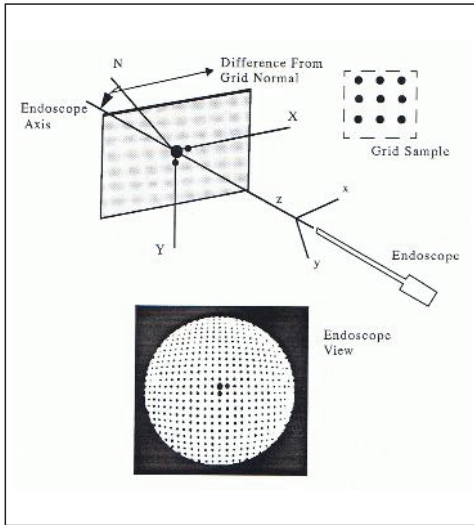

Figure 1. Endoscope camera calibration

Registration tasks

Two registration problems should be solved: (1) registration between the spaces of motion tracking device and diagnostic volume (CT, MR image sequence); (2) registration of video image to 3D space of motion tracking device. Next the procedure for video image registration is described.

Video frame registration

The registration matrix between the motion tracking device and the coordinate system of undistorted video $(\boldsymbol{T}_{W,C})$ can be given as

$$[X_{C,i} \quad Y_{C,i} \quad Z_{C,i} \quad P_{C,i}] = [X_{W,i} \quad Y_{W,i} \quad Z_{W,i} \quad 1] \times \boldsymbol{T}_{W,C} \quad (15)$$

where $[X_{C,i} \quad Y_{C,i} \quad Z_{C,i} \quad P_{C,i}]$ represents the coordinates within the perspective camera space and $[X_{W,i} \quad Y_{W,i} \quad Z_{W,i} \quad 1]$ is the position in the tracker's space. The matrix elements with nonzero last column values according to perspective projection:

$$\boldsymbol{T}_{W,C} = \begin{bmatrix} T_{1,1} & T_{1,2} & T_{1,3} & T_{1,4} \\ T_{2,1} & T_{2,2} & T_{2,3} & T_{2,4} \\ T_{3,1} & T_{3,2} & T_{3,3} & T_{3,4} \\ T_{4,1} & T_{4,2} & T_{4,3} & T_{4,4} \end{bmatrix} \quad (16)$$

The perspective division should be present in calculation of real image coordinates:

$$[X_{C,i}^* \quad Y_{C,i}^* \quad Z_{C,i}^* \quad 1] = [X_{C,i}/P_{C,i} \quad Y_{C,i}/P_{C,i} \quad Z_{C,i}/P_{C,i} \quad 1] \quad (17)$$

These equations can be rearranged if :

$$X_{C,i}^* \frac{(X_{W,i}T_{1,4} + Y_{W,i}T_{2,4} + Z_{W,i}T_{3,4} + 1)}{X_{W,i}T_{1,1} + Y_{W,i}T_{2,1} + Z_{W,i}T_{3,1} + T_{4,1}} = 1 \quad (18)$$

$$Y_{C,i}^* \frac{(X_{W,i}T_{1,4} + Y_{W,i}T_{2,4} + Z_{W,i}T_{3,4} + 1)}{X_{W,i}T_{1,2} + Y_{W,i}T_{2,2} + Z_{W,i}T_{3,2} + T_{4,2}} = 1 \quad (19)$$

from

$$X_{C,i}^* = X_{W,i}T_{1,1} + Y_{W,i}T_{2,1} + Z_{W,i}T_{3,1} + T_{4,1} - \quad (20)$$
$$- (X_{C,i}^* X_{W,i}T_{1,4} + X_{C,i}^* Y_{W,i}T_{2,4} + X_{C,i}^* Z_{W,i}T_{3,4})$$

$$Y_{C,i}^* = X_{W,i}T_{1,2} + Y_{W,i}T_{2,2} + Z_{W,i}T_{3,2} + T_{4,2} - \quad (21)$$
$$- (Y_{C,i}^* X_{W,i}T_{1,4} + Y_{C,i}^* Y_{W,i}T_{2,4} + Y_{C,i}^* Z_{W,i}T_{3,4})$$

this can be written into matrix equation with 11 unknowns $\boldsymbol{A} \times \boldsymbol{x} = \boldsymbol{b}$ $(T_{4,4} = 1)$:

$$A = \begin{bmatrix}
X_{W,1} & 0 & -X^*_{C,1}X_{W,1} & Y_{W,1} & 0 & -X^*_{C,1}Y_{W,1} & Z_{W,1} & 0 & -X^*_{C,1}Z_{W,1} & 1 & 0 \\
0 & X_{W,1} & -Y^*_{C,1}X_{W,1} & 0 & Y_{W,1} & -Y^*_{C,1}Y_{W,1} & 0 & Z_{W,1} & -Y^*_{C,1}Z_{W,1} & 0 & 1 \\
X_{W,2} & 0 & -X^*_{C,2}X_{W,2} & Y_{W,2} & 0 & -X^*_{C,2}Y_{W,2} & Z_{W,2} & 0 & -X^*_{C,2}Z_{W,2} & 1 & 0 \\
0 & X_{W,2} & -Y^*_{C,2}X_{W,2} & 0 & Y_{W,2} & -Y^*_{C,2}Y_{W,2} & 0 & Z_{W,2} & -Y^*_{C,2}Z_{W,2} & 0 & 1 \\
X_{W,3} & 0 & -X^*_{C,3}X_{W,3} & Y_{W,3} & 0 & -X^*_{C,3}Y_{W,3} & Z_{W,3} & 0 & -X^*_{C,3}Z_{W,3} & 1 & 0 \\
0 & X_{W,3} & -Y^*_{C,3}X_{W,3} & 0 & Y_{W,3} & -Y^*_{C,3}Y_{W,3} & 0 & Z_{W,3} & -Y^*_{C,3}Z_{W,3} & 0 & 1 \\
X_{W,4} & 0 & -X^*_{C,4}X_{W,4} & Y_{W,4} & 0 & -X^*_{C,4}Y_{W,4} & Z_{W,4} & 0 & -X^*_{C,4}Z_{W,4} & 1 & 0 \\
0 & X_{W,4} & -Y^*_{C,4}X_{W,4} & 0 & Y_{W,4} & -Y^*_{C,4}Y_{W,4} & 0 & Z_{W,4} & -Y^*_{C,4}Z_{W,4} & 0 & 1 \\
X_{W,5} & 0 & -X^*_{C,5}X_{W,5} & Y_{W,5} & 0 & -X^*_{C,5}Y_{W,5} & Z_{W,5} & 0 & -X^*_{C,5}Z_{W,5} & 1 & 0 \\
0 & X_{W,5} & -Y^*_{C,5}X_{W,5} & 0 & Y_{W,5} & -Y^*_{C,5}Y_{W,5} & 0 & Z_{W,5} & -Y^*_{C,5}Z_{W,5} & 0 & 1 \\
X_{W,6} & 0 & -X^*_{C,6}X_{W,6} & Y_{W,6} & 0 & -X^*_{C,6}Y_{W,6} & Z_{W,6} & 0 & -X^*_{C,6}Z_{W,6} & 1 & 0
\end{bmatrix} \tag{22}$$

$$x = \begin{bmatrix} T_{1,1} \\ T_{1,2} \\ T_{1,4} \\ T_{2,1} \\ T_{2,2} \\ T_{2,4} \\ T_{3,1} \\ T_{3,2} \\ T_{3,4} \\ T_{4,1} \\ T_{4,2} \end{bmatrix} \qquad b = \begin{bmatrix} X^*_{C,1} \\ Y^*_{C,1} \\ X^*_{C,2} \\ Y^*_{C,2} \\ X^*_{C,3} \\ Y^*_{C,3} \\ X^*_{C,4} \\ Y^*_{C,4} \\ X^*_{C,5} \\ Y^*_{C,5} \\ X^*_{C,6} \end{bmatrix} \tag{23}$$

The solution vector *(x)* can be calculated (for quadratic matrix in (22)) by simple matrix inversion (originally DLT method[13], for X-Ray processing see[14]). Unfortunately, this doesn't work in practice (especially in video image registration) because of marker localization errors. Statistical approach (with SVD algorithm[11]) was found much more promising[1,2] which is applicable for over-determined equations of more than 6 markers (acceptable accuracy is possible with 7-8 markers). Marker localization is possible by detecting the tip of the probe of tracking device in "frame-freezing" mode. In case of endoscopes, the "undistorter" step (see Video calibration section) provides for linearized camera position which is suitable for 3D registration.

Results

For this kind of complex project, establishing application concepts and suitable architecture of system elements (hardware, software) appear as major tasks. The results at this time are technical and highly depend on resources available for development. Important elements of clinical application[8], which have been made
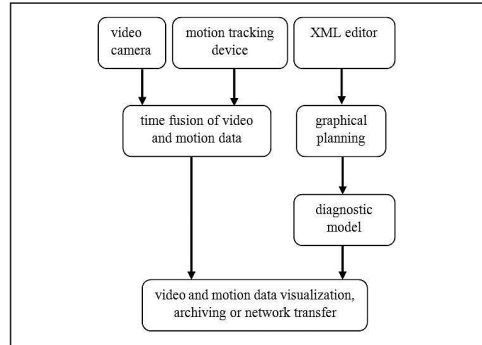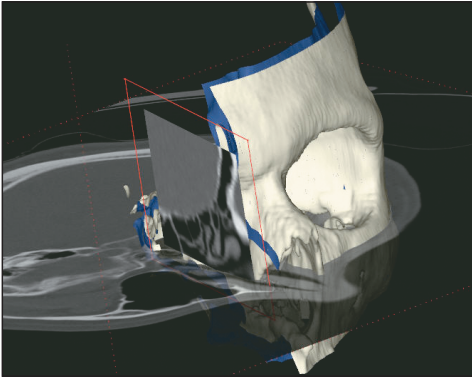

Figure 2. Simplified system architecture

already available or developed by own resources, are: (1) video digitalization with DT3131 frame-grabber board (Data Translation Inc., USA) *(http://www.datatranslation.com/products/imaging/pci/);* (2) integrating GPU (graphical processing unit) *(http://www.nvidia.com)* with GLSL programming („Open GL Shading Language" *http://www.opengl.org/documentation/glsl/*); (3) use of modern 6 DOF mouse in surgical planning *(http://www.3Dconnexion.com);* (4) software control of optical motion tracking for different Northern Digital cameras (Polaris Standard and Vicra, http://www.ndigital.com); (5) use of XML (Extensible Markup Language, *http://www.w3.org/XML/)* compressed files for archiving, editing and transmitting information on surgical plan, registration and camera distortion parameters, tool calibration, diagnostic data, etc. XML files guarantee the easy communication through internet and are important parts of telemedicine applications.

Figure 3. Use of modern GPU in modeling diagnostic data. The calculation of arbitrarily oriented cutting planes from a skull CT and comparison to surface graphics occurs in real-time according to actual state of the 6 DOF mouse

Figure 2 represents the core elements of the system which is under investigation. *Figure 3* illustrates the practical advantage of GPU programming with GLSL language.

Discussion

This article gives an introduction to a new project which can be considered as the author's personal initiative based on earlier experience. Hopefully, in the future, it can gain more institutional support especially for software development. In this case there is a chance to move faster with new applications of real clinical value and usability. Procedure, similar to the presented one in this paper, for tracking the location of endoscopic image relative to CT volume is described in[16] but without time-synchronized archiving of video and motion data. Interesting new approach is already known to integrate endoscopic video with CT diagnostic data[15] without motion sensor attached to endoscope. However, this direct (based on image feature detection) approach is not suitable for clinical use yet.

Valuable survey has been given about human motion tracking systems[3]. The classification distinguishes vision based (without using motion tracking device) and non-vision based (with motion tracking device) applications. The fusion of video and tracking data, as proposed in this article, usually is available in expensive, commercial systems (Elite Biomech, Vicon) but without clear information on technology behind. The need for subsequent, repeated analysis or network transfer of motion data certainly demands fused, compressed representation of results with video sequence and kinematical models. This paper outlines also the involvement of preoperative diagnostic images in motion analysis and an effective control of the complex environment through XML file representation.

Conclusions

Overview and mathematical description of most critical parts of fusing video and motion data have been given. The results can be used in two areas: (1) for training and reliability improvement of endoscopic surgeries of high risk (like in ENT); (2) in kinematical modeling of human movement by representing video image and graphical models in a common reference space.

In Endoscopic Paranasal Sinus surgeries[7,8] and other fields of endoscopic intervention the real fusion (in time and 3D space) of video sequence and tools' position makes archiving or transmitting information on actual state of surgery possible. This can be used for evaluating problematic results or educating risky interventions.

In biomechanical research, the proposed approach solves the critical problem of human motion analysis if soft tissue markers can't be used for bone axis definition or diagnostic

images (X-Ray, CT or MR) are not available. However, if valid registration between body parts and diagnostic images is available, the video-based representation of bone axis can be projected back and compared to the static diagnostic volume during motion.

REFERENCES

1. *Truppe M, Pongracz F, Ploder O, Wagner A* and *Ewers R*. Interventional Video Tomography. In Proceedings of Lasers in Surgery, vol. 2395. San Jose, CA: SPIE 1995:150–152.

2. *Freysinger W, Truppe M, Gunkel AR, Thumfart WL, Pongracz F,* and *Maierbaeuerl J*. Interactive telepresence and augmented reality in ENT surgery : Interventional Video Tomography. In: Proceedings of CVRmed – MRCAS 97, Grenoble, 1997;1205:817–820.

3. *Zhou H, Hu H*. Human motion tracking for rehabilitation–A survey. Biomedical Signal Processing and Control 2008 Jan;3(1):1–18.

4. *Amaya K, Hara Y, Aoki S*. Reconstruction of 3D human movement using inverse analysis. Inverse Problems in Engineering Mechanics. In: Proceedings of International Symposium on Inverse Problems in Engineering Mechanics (ISIP'98) 1998; Nagano, Japan, 1998:583–592.

5. *Luo Y, Lopez FJP* and *Pipaon JJV*. An automatic rotoscopy system for human motion based on a biomechanic graphical model. Computers & Graphics 1992;16(4):355–362.

6. *Berkelman P, Cinquin P, Boidard E, Troccaz J, Létoublon C, Ayoubi JM*. Design, control and testing of a novel compact laparoscopic endoscope manipulator. Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering 2003;217(4): 329–341.

7. *Kumagai T. et al*. Distance Education System for Teaching Manual Skills in Endoscopic Paranasal Sinus Surgery Using "HyperMirror" Telecommunication Interface. In: Proceedings of IEEE Virtual Reality, 2008, March 8–12, Reno, Nevada: 233–236.

8. *Pongrácz F*. Use of video motion tracking and telepresence for training and reliability improvement of endonasal surgery. Supported project by National Office for Research and Technology (NKTH), Budapest 2009 under OMFB-01134/2009.

9. *Marin F, Mannel H, Claes L* and *Dürselen L*. Correction of axis misalignment in the analysis of knee rotations. Human Movement Science 2003 Aug;22(3):285–296.

10. *Bowman ME* and *Forrest AK*. Transformation calibration of a camera mounted on a robot. Image and Vision Computing 1987;5(4):261–266.

11. *Press WH, Teukolsky SA, Vetterling WT, Flannery BP*. Numerical Recipes in C. The Art of Scientific Computing. Cambridge: Cambridge University Press; 1992.

12. *Andrónyi K* and *Szobonya L*. The importance of axis alignment in the human knee examination. In: Proceedings of Third Hungarian Conference on Biomechanics, 2008, July 4–5, Budapest: 13–22.

13. *Abdel-Aziz YI* and *Karara HM*. Direct linear transformation into object space coordinates in close-range photogrammetry. In: Proceedings of the Symposium on Close-Range Photogrammetry, 1971, Jan. University of Illinois, 1971:1–8.

14. *Fekete K et al*. X-Ray image processing by direct linear transformation. In: Proceedings of Third Hungarian Conference on Biomechanics, 2008, July 4–5, Budapest; 75–81.

15. *Mirota D et al.* Toward video-based navigation for endoscopic endonasal skull base surgery. In: Proceedings of MICCAI 2009, Lecture Notes in Computer Science, Springer Berlin/ Heidelberg, Vol. 5761/2009:91–99.

16. *Shahidi et al.* Implementation, calibration and accuracy testing of an image-enhanced endoscopy system. IEEE Trans Med Imaging 2002; 21(12):1524–35.

Ferenc Pongrácz
Project Employment Supported by 5LET 2008 Contract with
Bureau of National Research and Development in Hungary (NKTH)
Phone: (+36) 1 275-8615
Mobile: (+36) 30 562-1806