

Surgical Training System with a Novel Approach to Human-Computer Interaction

A. Wytyczak-Partyka¹, J. Nikodem¹, R. Klempous¹, and J. Rozenblit²

¹ Wrocław University of Technology, Wrocław

ryszaard.klempous@pwr.wroc.pl

² University of Arizona, Tucson

jr@ece.arizona.edu

Abstract. In this chapter we present a surgical training system that employs some novel ideas on interaction between the trainee and the system.

1 Introduction

Laparoscopic surgery brings significant benefits into the healing process and therefore an increasing interest in it is observed. Great benefits over traditional surgery include: limited scarring, reduction in pain and recovery time, leading to a smaller risk of complications. Study conducted by Hansen et al. [1] shows that patients who have undergone laparoscopic appendectomy had five times fewer wound infections, two times shorter discharge time and fewer of them required narcotic analgesia. On the other hand, there is a number of downsides, for instance the surgeon's perception – both haptic and visual – is very limited, which extends the procedure time (in the open appendectomy case 40 as opposed to 63 minutes in laparoscopic appendectomy [1]) and the likelihood of human error. Also investment in expensive instruments and a very long training period are required.

Surgical training should be modular, where each module should focus on the development of certain behaviors – knowledge-based, rule-based and skill-based. This chapter will focus on a system designed particularly for the skill-based behavior training.

The primary skills that have to be developed during the training involve – depth perception, dexterity and hand-eye coordination. Traditionally, according to the Surgical Papyrus [2] since the times of ancient Egypt, surgical training has always followed the model of master and apprentice and involved mentorship in real-life clinical cases, where the apprentice would gain skills and experience from his teacher. That approach to surgical education hasn't changed significantly since ancient times. It is also worth noting that the training was strictly dependent on the availability of a patient and tied with the course of patient care [3-5].

Since some of the surgical skills do not strictly require practicing them in a real-life clinical situation – it is desirable to enhance them in a safe environment, without any risk to patients. It is also important to note, that the operating room is not the best learning environment because of factors such as stress, time constraints and costs that have a negative impact on the learning process.

On the other hand, reference [6] shows, that surgical simulation has positive impact on improvement of psychomotoric skills, i.e. the gallbladder dissection procedure was 29% faster for residents trained on simulators, also accidental injuries during the procedure were 5 times less frequent in that group. Similar results in a cholecystectomy procedure are shown in [7]. Therefore a number of surgical simulators, both physical-model-based and software-based, have been developed, all allowing the trainees to safely master the basics skillset.

One of the first simulation devices for laparoscopic training was the Pelvi-trainer, designed by Karl Semm in 1986 [8]. The original concept consisted of a transparent box, where organs were put. The box contained several holes to introduce the instruments and the camera, Figure 1. The training method proposed by Semm was gradual: at first the trainee would only learn to use the instruments, without the use of the endoscope, secondly the endoscope would be introduced, while the trainee would be still allowed to occasionally look at the organ through the transparent walls of the box, and finally the box would be covered with a cloth to obscure vision.

Since 1986 the concept of a pelvi-trainer has influenced many simulators and is currently used in many training programs.

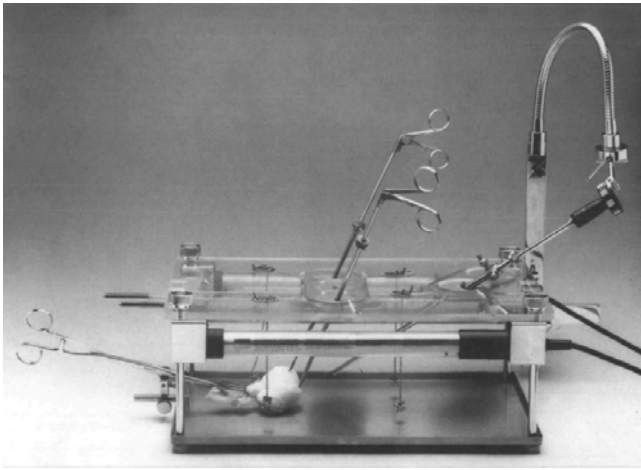


Fig. 1. Karl Semm's pelvi-trainer [8]

Semm's pelvi-trainer is a typical example of a skill-based training device that focuses on development of dexterity, depth perception and hand-eye coordination. There are several simulators that serve for knowledge- and/or behavior-based training as well as skill-based.

One of the first systems built for that purpose is the Karlsruhe Endoscopic Surgery Trainer and the KISMET software [9, 10]. It is a complete training system with built in scenarios for practicing procedures like laparoscopic cholecystectomy. There are several products similar to the Karlsruhe Trainer, for instance the MISTELS system [11], or the LapMentor [12]. A lot of effort has been put in

those systems to reproduce details such as graphics, organ deformations and haptic feedback. Needless to say those systems are very expensive.

One skill-based approach that combines a simple pelvi-trainer with a high-tech system is the Virtual Assistant Surgical Training (VAST) system, developed at the University of Arizona [13]. It is comprised of a pelvi-trainer and a computer system. The computer, through a magnetic position sensor, collects data about the instrument's tip position, which is used to rate the trainee's performance in a certain exercise, based on time, path length and accuracy. The individual's progress can therefore be precisely measured and monitored.

The approach represented by VAST is especially interesting and will serve as a basis for the further described training system which incrementally adds to the VAST trainer.

2 Description of the System

The purpose of the proposed system is to aid a trainee in the development of basic skills, as dexterity, hand-eye coordination and depth perception. The novel way of interaction proposed in this chapter – allows the trainee to develop the skill of avoiding certain regions where the appearance of an instrument might inflict damage to the patient.

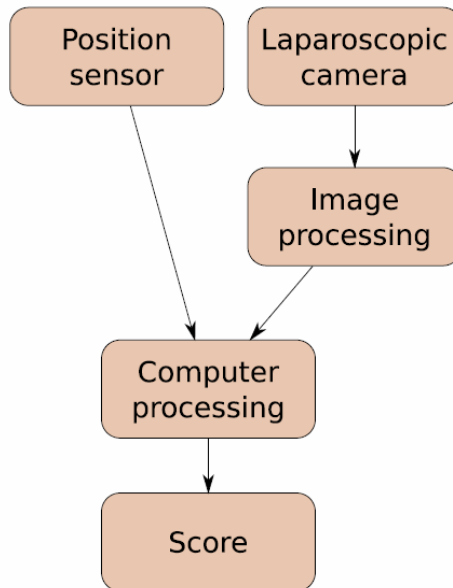


Fig. 2. System outline

2.1 System Outline

The system resembles the VAST prototype trainer [13] and is comprised of a standard pelvi-trainer setting – an endoscopic camera and two instruments - and a

computer. The instruments have an embedded position sensor. The video output of the camera, as well as the position sensors, are connected to the computer, where additional processing occurs, as in Figure 2. There are two results of the processing. First - each exercise performed by a trainee is scored, and thus the performance can be analyzed, and secondly – the trainee is optionally informed by an auditory signal coming from a speaker, whenever he approaches the hazardous area with the instruments.

The score achieved in the exercises is calculated by the computer and based on the information collected from position sensors and the video camera.

Initially, before the exercise can start, and after a new model is introduced to the pelvi-trainer – a 3D representation of the model is built, from images collected with the endoscopic camera. Secondly – a hazardous region can be selected, the trainee has the opportunity to see and understand where the region is on the model, so it can later be avoided, during the course of the exercise. This can be done individually for each exercise to introduce more difficulty into the exercises.

Later the coordinates of the instruments obtained from the position sensors can be used to determine the location of the instrument in the reconstructed environment.

The 3D structure of the viewed scene, recovered in the process of the system's work, is at this point not intended strictly for visualization - the trainee observes the image as it comes from the endoscopic camera, but rather for the purpose of determining the hazardous regions within the 3D model. Through fusion of the 3D reconstructed model and sensor data the coordinates collected from the position sensor located on the tip of the instrument can be used to determine its location and proximity to the hazardous area. If the position sensor indicates appearance of the instrument in a hazardous region an auditory signal is produced and a penalty score applied.

2.2 Training Model

The training is based on simple tasks that trainees have to complete within a specified time. Several types of exercises have been proposed by others [14, 15], the common goal is to practice dexterity, coordination and depth perception.

Example exercises are:

- knot tying,
- cutting and suturing,
- picking up objects,
- touching different points on a model.

Each exercise is associated with a different physical model which is placed in the pelvi-trainer's box.

The trainee's score is calculated with respect to:

- elapsed time,
- length of the path of the instrument tip,
- accuracy.

It is proposed that another factor is introduced and that the score is significantly decreased upon hitting a hazardous region, defined at the beginning of each exercise, therefore the scoring function changes from

$$S = f\left(\frac{k_t}{t}, \frac{k_s}{s}, k_A \cdot A\right), \quad (1)$$

to

$$S = f\left(\frac{k_t}{t} + \frac{k_s}{s} + k_A \cdot A + k_H \cdot H\right), \quad (2)$$

with H , the hazard measure, defined as

$$H = -\frac{2R_H}{|d(H_c, T)| + R_H} + 1 \quad (3)$$

The terms H_c , T and R_H in equation 3 denote respectively - the coordinates of the center of the hazard sphere, the coordinates of the instrument's tip and the radius of the hazard sphere.

Such a way of scoring the trainee allows considering the events of breaching the no-fly zones and is an important part of the training part of the system.

2.3 Processing

The following section describes the method for recovering the 3D structure of the scene applied in the presented system, which is the initial step in the work of the proposed system.

During laparoscopic procedures it is natural that after introducing the camera into the body, the surgeon performs a series of movements with the camera to find the best viewing point for performing the procedure, and also to discover any potential abnormalities in the viewed organs. It is therefore natural that a video sequence containing images of the operating field can be obtained in that manner also during the training procedure and serve as a source of images for a structure from motion 3D geometry recovery algorithm.

Structure from Motion

Since the early 1980's there has been a lot of research in the field of recovering 3D structure from camera image sequences [16, 17]. Current state of the art algorithms [18, 19] perform the task without previous camera calibration and with the only constraints on the image sequence that a certain number of corresponding feature points between frames can be established and that a sufficient baseline (distance between the camera locations) exists. It has been proven [18] that the point-correspondences, along with several constraints on the camera, can lead to a metric reconstruction of the imaged scene.

In general the reconstruction process consists of: (a) selecting potential point matches between the frames of the sequence, (b) estimating the 3D geometry, (c) refinement. Since the whole process is based on point correspondences it is very important that the first step is performed carefully. Therefore potential matches, (m, m') , are selected from a set of points, populated through a Harris' corner detection procedure [20], that contains only interest points that significantly differentiate from their neighborhood and can be matched with their corresponding points in other views using a similarity measure, i.e. cross-correlation. It is important that the usage of a video sequence limits the search for matches to a certain subwindow of the image, because the camera movement between the frames (baseline) is small and therefore feature points stay in a several-pixel range.

After the feature matching has concluded two initial views are selected that will serve as a base for further sequential structure refinement. The criteria of selecting such views are: a) maximization of the number of features matched between the views, and b) sufficient baseline between the views, so the initial 3D estimate can be properly performed.

While the first criterion is easy to fulfill, based on the results of feature matching, it is significantly harder to determine if a view meets the second one, which is especially important in images from a video sequence, where the baselines are usually very small.

The simplest approach would be to limit the frame sequence and select only every k -th frame, where k depends on the frame rate of the video. In fact, basing on the constraint that the camera moves are rather smooth and slow and the imaged object itself remains still, it is suggested to perform such a reduction of the input video sequence, by reducing the frame rate to 2 frames per second. Such a reduction essentially decreases the computational effort involved with the feature point selection.

However a robust criterion for selecting key-frames is available [21] and has been employed in the proposed system.

The GRIC, Geometric Robust Information Criterion [21], criterion can be used to determine if the geometry between 2 views is better explained by a homography transformation (true for small baselines) or by the fundamental matrix (larger baselines). Therefore GRIC is suitable to select views which are interesting for the 2-view reconstruction process which only deals with the fundamental matrix model. It is especially important to evaluate GRIC for image sequences obtained with a video camera, where baselines can be really small for consecutive frames, it is less significant in case of image sequences from a still-camera, where most of the baselines are wide enough and the fundamental matrix model scores better anyway - as seen in Figure 4.3(a).

GRIC is a scoring function that takes into account several factors - number n of inliers plus outliers, the residuals e_i , standard deviation of measurement error, the dimension of the data r (4 for two views), number k of motion parameters ($k = 7$ for F, $k = 8$ for H) and the dimension d of the structure ($d = 3$ for F, $d = 2$ for H).

$$GRIC = \sum \rho(e_i^2) + (nd \ln(r) + k \ln(rn)). \quad (4)$$

After the initial two views have been selected the images may be further processed for the retrieval of the 3D geometry, next additional views are added, selected in a similar manner as in the initial step. The recovery of the 3D geometry is based on the normalized 8-point algorithm and a triangulation method described in [19]. The 8-point algorithm is used to compute the fundamental matrix, F , and the epipoles (e, e'), from a set of point correspondences (m, m') between 2 images by solving (5).

$$m'^T F m = 0. \quad (5)$$

The F matrix can be then used to calculate the camera matrices, (P, P'), which are needed for the triangulation step.

$$\begin{aligned} P &= [I_{3 \times 3} | 0], \\ P' &= [e']_x F [e'], \end{aligned} \quad (6)$$

where $[e]_x$ is a skew symmetric matrix of the form:

$$[a]_x = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}, \quad (7)$$

and linked with vector product in the following manner:

$$a \times b = [a]_x b. \quad (8)$$

The F matrix is calculated using the RANSAC algorithm, proposed by Fischler and Bolles [22], which is less sensitive to outliers in the matched point set than the standard 8-point algorithms, however it has to be fed with a greater number of potential matches.

The information obtained so far is sufficient to calculate rectifying transformations for the image pairs, which further serve as input for a stereo depth matching algorithm that produces a dense depth map of the scene.

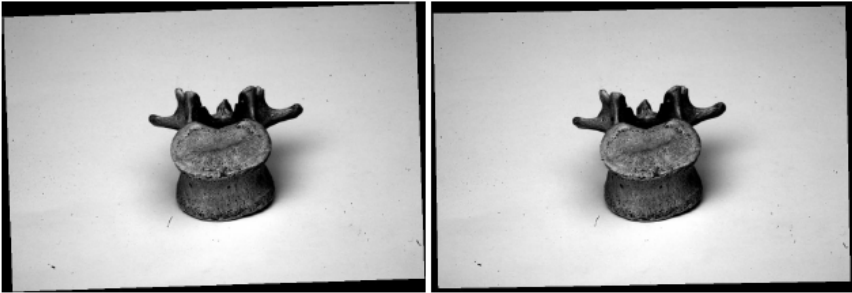


Fig. 3. Rectified image pair

A triangular mesh of the scene is constructed using the Delaunay [23] triangulation method. The mesh is used for visualization of the 3D model of the operating field where at this point no-fly zones can be oriented.

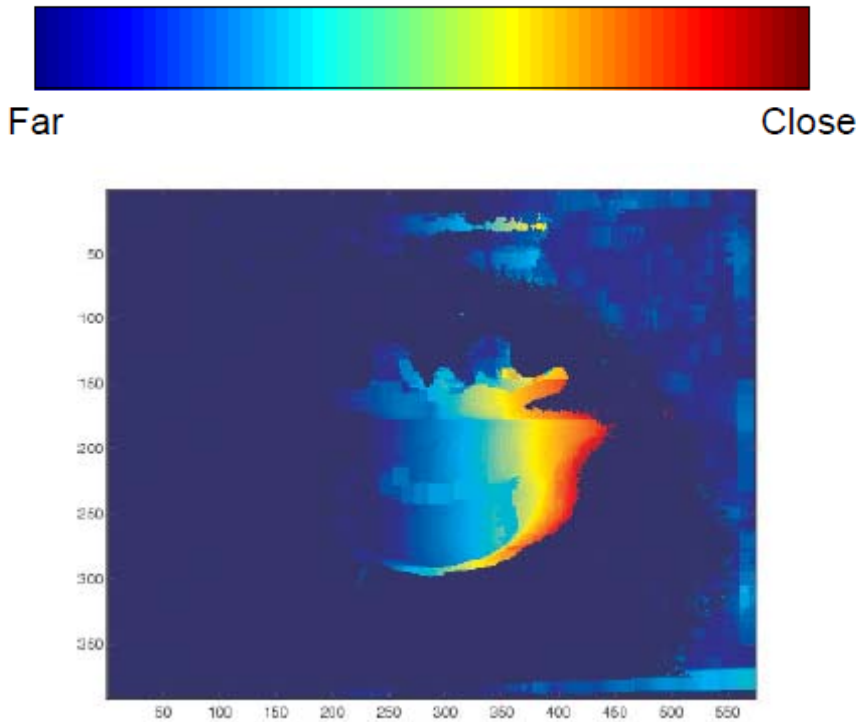


Fig. 4. Non-interpolated depth map of the vertebra, obtained with a segmentation-based depth matching algorithm

Interaction

The interactions are based on the information about the trainee's behavior with respect to the location of no-fly zones. By using fuzzy classification of movements [24] the system can guide the trainee to obtain better results. The advice offered by the system is of the form:

- your movements are too fast,
- your movements are too imprecise,
- your moving too close to a no-fly zone,

also the system can advise the user to go to a certain exercise which is designed to emphasize the particular skill the system considers to be weak.

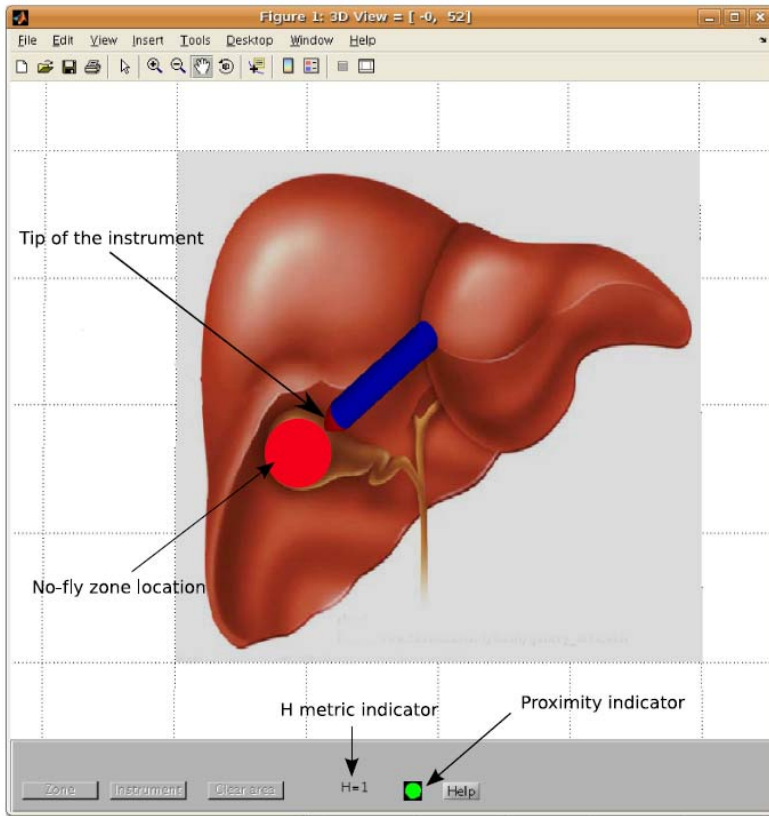


Fig. 5. Illustration of the concept of the simulator of the system, showing example of no-fly zone location and the interaction features like the no-fly zone proximity indicator

3 Conclusions

This chapter describes a training system which utilizes a new interaction scheme for the purpose of aiding in the process of training laparoscopic surgeons.

The video sequence obtained from the laparoscopic camera is used to construct a 3D model of the operating field in which information about no-fly zones is embedded. The locations of no-fly zones and the position of the instrument tip are then used to provide the interaction and guidance during training.

Knowledge of the 3D model and geometry of the scene can be further used to visually augment additional information into video from the endoscopic camera. It is interesting to apply the augmented-reality approach to laparoscopic training and examine its usability in the Operating Rooms. The presented system can serve as a basis for such augmented-reality trainer.

References

1. Hansen, J.: Laparoscopic versus open appendectomy: prospective randomized trial. *World J. Surg.* 20(1), 17–21 (1999)
2. Breasted, J.: *The Edwin Smith surgical papyrus*. University of Chicago Press, Chicago (1991)
3. Gorman, P., Meier, A., Krummel, T.: Computer-assisted training and learning in surgery. *Comput Aided Surg.* 5, 120–130 (2000)
4. Cosman, P., Cregan, P., Martin, C., Cartmill, J.: Virtual reality simulators: current status in acquisition and assessment of surgical skills. *ANZ J. Surg.* 72, 30–34 (2002)
5. Kneebone, R.: Simulation in surgical training: educational issues and practical implications. *Med. Educ.* 37, 267–277 (2003)
6. Seymour, N., Gallagher, A., Roman, S., OBrien, M., Bansal, V., Andersen, D., Satava, R.: Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann. Surg.* 236, 458–463 (2002)
7. Grantcharov, T., Kristiansen, V., Bendix, J., Bardram, L., Rosenberg, J., Funch-Jensen, P.: Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *Brit. J. Surg.* 91, 146–150 (2004)
8. Semm, K.: Pelvi-trainer, a training device in operative pelviscopy for teaching endoscopic ligation and suture technics. *Geburtshilfe Frauenheilkund* 46, 60–62 (1986)
9. Kuhnappel, U., Akmak, H., Maaß, H.: 3D Modeling for endoscopic surgery. *Proc. IEEE SIMSYS*, 22–32 (1999)
10. Kuhnappel, U., Krumm, H., Kuhn, C., Hubner, M., Neisius, B.: Endosurgery simulations with KISMET: a flexible tool for surgical instrument design, operation room planning and VR technology based abdominal surgery training. *Proc. IEEE VR* 95, 165–171 (1995)
11. Fraser, S., Klassen, D., Feldman, L., Ghitulescu, G., Stanbridge, D., Fried, G.: Evaluating laparoscopic skills. *Surg. Laparosc. Endosc.* 17, 964–967 (2003)
12. Simbionix Web Site Laparoscopy Simulators, <http://www.simbionix.com> (accessed February 22, 2007)
13. Feng, C., Rozenblit, J., Hamilton, A.: A hybrid view in a laparoscopic surgery training system. *Proc. IEEE ECBS*, 339–348 (2007)
14. Derossis, A., Fried, G., Abrahamowicz, M., Sigman, H., Barkun, J., Meakins, J.: Development of a model for training and evaluation of laparoscopic skills. *Am. J. Surg.* 175, 482–489 (1998)
15. Fried, G., Feldman, L., Vassiliou, V., Fraser, S., Stanbridge, D., Ghitulescu, G., Andrew, C.: Proving the value of simulation in laparoscopic surgery. *Ann. Surg.* 240, 518–528 (2004)
16. Horn, B., Schunck, B.: Determining Optical Flow. *Artificial Intelligence* 17, 185–203 (1981)
17. Ullman, S., Hildreth, E.: The measurement of visual motion. In: Braddick, O., Sleight, A. (eds.) *Physical and biological processing of images*. Springer, Berlin (1983)
18. Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., Koch, R.: Visual modeling with a hand-held camera. *Intern. J. Computer Vision* 59, 207–232 (2004)
19. Hartley, R., Zisserman, A.: *Multiple view geometry in computer vision*. Cambridge University Press, Cambridge (2003)

20. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proc. Alvey Vision Conference, pp. 189–192 (1988)
21. Torr, P., Fitzgibbon, A., Zisserman, A.: The problem of degeneracy in structure and motion recovery from uncalibrated Image Sequences. *Intern. J. Comput. Vision* 32, 27–44 (1999)
22. Fischler, M., Bolles, R.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* 24, 381–395 (1981)
23. Delaunay, B.: Sur la sphere vide. *Izv AN SSSR. Otd Mat. Estest Nauk* 7, 793–800 (1934)
24. Wytyczak-Partyka, A., Nikodem, J., Klempous, R., Rozenblit, J., Feng, C.: Computer-guided laparoscopic training with application of a fuzzy expert system. In: Gelbukh, A., Morales, E.F. (eds.) *MICAI 2008. LNCS (LNAI)*, vol. 5317, pp. 965–972. Springer, Heidelberg (2008)