Motion Detection System with GPU Acceleration for Stereotactic Radiosurgery

Takuya YAMAKAWA*1, Koichi OGAWA*1, Hitoshi IYATOMI*1, Etsuo KUNIEDA*2, Keisuke USUI*1,2, Naoyuki SHIGEMATSU*3

Abstract

Stereotactic radiosurgery is a non-invasive method for the treatment of tumors that employs a narrow, high-energy X-ray beam. In this form of therapy, the target region is intensively irradiated with the narrow beam, and any unexpected patient motion may therefore lead to undesirable irradiation of neighboring normal tissues and organs. To overcome this problem, we propose a contactless motion detection system with three USB cameras for use in stereotactic radiosurgery of the head and neck. In our system, the three cameras monitor images of the patient’s nose and ears, and patient motion is detected using a template-matching method. If patient motion is detected, the system alerts the radiologist to turn off the beam. We reduced the effects of variations in the lighting in the irradiation room by employing USB cameras sensitive to infrared light. To detect movement in the acquired images, we use a template-matching method that is realized with general-purpose computing-on-graphics processing units. In this paper, we present an outline of our proposed motion detection system based on monitoring of images of the patient acquired with infrared USB cameras and a template-matching method. The performance of the system was evaluated under the same conditions as those used in actual radiation therapy of the head and neck.

Key words: Stereotactic radiosurgery, Motion detection, Template matching, Infrared light, GPU, USB camera

Med Imag Tech 30(5): 268-278, 2012

1. Introduction

Stereotactic radiosurgery is one of the noninvasive treatment methods in which a targeted tumor is irradiated precisely with a narrow high-energy x-ray beam. The therapeutic gain of such a noninvasive method is compatible with surgical operations, and so this radiosurgery is gradually being increasingly applied. In this therapy, a target region is intensively irradiated with a narrow beam, and thus any unexpected patient motion may lead to undesirable irradiation of neighboring normal tissues and organs. In radiation therapy of the head and neck region, many frames such as the Leksell frame have been developed [1] for fixing the patient’s head tightly to the couch. However, most such frames are cumbersome to use and painful for the patient. For solving this problem, several methods for tracking patient or tumor motion have been proposed.

Target tracking with a fluoroscopic imaging system [2–4] uses two x-ray images of a target region viewed from different angles. This system uses gold markers inserted near the target position to visualize the target position clearly [2, 3]. A marker-less method has also been proposed [4]. These methods facilitate high accuracy in radiation therapy, but the fluoroscopic imaging that is used for checking the target position increases the dose to normal tissues around the target; in addition, the system is relatively expensive. The optical tracking methods [3, 5, 6] use
infrared markers located on the surface of a patient’s body. These markers are detected with two charge-coupled-
device (CCD) cameras, and the three dimensional locations of the markers are calculated with parallax information.
These methods can obtain accurate positions of the markers, although the movement of these markers and that of
the target do not always correspond, and the detection system is expensive. Tracking systems with electromagnetic
transponders sealed in a glass container (8 mm in length × 1.85 mm in diameter) have also been proposed [7~9]. In
these, three electromagnetic transponders are inserted near the target organ, and the electromagnetic signal (10 Hz)
emitted from the transponders is used for detecting the movement of the target organ. This system is also used in
combination with the On-board-imaging system described by Santanam et al [9]. The accuracy of position detection
of the system is sub-millimeter, but the system requires implantation of these transponders, which is invasive
for the patient.
We also studied a motion detection system for several years and proposed a contactless system [10, 11]. In our
system, the movement of the patient’s head was detected by means of a template-matching method with three USB
(universal serial bus) cameras located around the patient’s head. This system uses a commonly available PC and
USB cameras; therefore, it is not expensive, and we need not be concerned about excessive irradiation or invasive
procedures. However, our previous system did not adequately deal with changes in the illuminance induced by rotation
of the linear accelerator gantry in the clinical setting. The template matching based on a normalized cross-cor-
relation [12, 13] or phase-only correlation [14] is effective in overcoming the above problem, although it is time-
consuming to calculate the correlation.
In our proposed method, we solved this problem by using USB cameras that are sensitive to infrared light, and we cut off the visible light with an infrared filter. To detect the patient’s movement, we adopted an active search algorithm [15] in our previous paper [10]. This search method calculates a similarity between a template and a search area with a histogram, and it can reduce the processing time significantly compared to that for an exhaustive search method. However, it takes a long time to calculate the similarity when we use three VGA images (640 × 480 pixels). On the other hand, recently a parallel-processing architecture such as OpenMP [16] has been available, and GPGPU (general-purpose computing on graphics processing units) with CUDA (compute unified device architecture) [17] has become another choice for dealing with this problem. However, the OpenMP is not ideal in terms of the calculation speed, and so we adopted the GPGPU instead of the OpenMP method to improve the detection speed of the patient’s movement with VGA images. In this paper, we outline our proposed system and show the results of experiments to confirm the accuracy of our motion detection system.

2. Materials and methods

1) Hardware configuration
In our proposed system, we used three USB cameras that detect the motion of some body parts of the patient, such as the ears and nose. We assumed that the targeted area in the radiotherapy was located in the brain. The position of the three USB cameras (DC-NCR20U, Hanwha Japan) is shown in Fig. 1; each camera was fixed to the camera arm located around the patient’s head. The energy of the x-ray beam was sufficiently high to damage these USB cameras, and therefore we positioned them as far as possible from the irradiation area of the target. We also used a bite plate which was attached to the camera arm to restrict the movement of the patient’s head. Fig. 2(a) shows a USB camera, and Fig. 2(b) and (c) show images of the nose, acquired under various lighting conditions.

The illuminance variation due to the rotation of the linear accelerator (linac) gantry makes it difficult to detect the movement of the target area. To reduce this influence, we used USB cameras that were sensitive to infrared light and illuminated the nose and two ears with the infrared light. These cameras were also sensitive the visible light; therefore, we cut off the visible light with an infrared filter (IR filter, Fujifilm, Tokyo, Japan), and acquired images of the infrared light. Fig. 2(d) shows the USB camera with the IR filter. In addition, we controlled the intensity of the infrared light with a thin paper filter. Fig. 2(e) and (f) show images of the nose acquired under different visible light conditions under illumination with infrared light.

The movement of the patient’s head was monitored with the three USB cameras connected to a personal com-
puter, and the acquired temporal images were processed with the template images acquired in the initial position of
the patient’s head. We acquired images with a resolution of 640 × 480 (VGA) with the above three USB cameras
connected to a PC (Precision M6500, [Core i7 920 Extreme edition 2 GHz, 4 cores, 16 GB memory], Dell, Texas,
USA), and the movement of the patient’s head was calculated with a GPU (Quadro FX3800M, [number of stream
processors: 128, 1 GB memory], NVIDIA, California, USA) to improve the image processing speed.

2) Software configuration
(1) Motion detection algorithm

Our method detects some body parts of the patient as marker objects in a pattern-matching manner. First, we cre-
ated a standard template for the specified object, such as an ear or nose, being trimmed from an acquired image
(Fig. 3). For each body part, motion tracking was performed on the acquired images during the radiotherapy with
each template image of an ear or nose. The most likely area between the template image and an acquired image for
each view was searched for with an exhaustive search method with the GPGPU. The GPU consisted of 16 multi-
processors (MPs). An MP consisted of eight stream processors (SPs) that calculate simultaneously. The term
‘block’ is a unit of parallel calculation in an MP, and ‘thread’ is a unit that an SP calculates simultaneously. In this
GPU, 128 threads (16 MPs × 8 SPs) calculate simultaneously. The GPGPU software was developed with use of
CUDA [17] (distributed by NVIDIA, California, USA). The CUDA can improve the calculation speed by using this
two-layered parallel processing.
The degree of matching was evaluated with the mean squared error (MSE). Fig. 4 shows the calculation of the MSE between the template image and the targeted area on a temporal frame by using a texture memory of the GPU. The texture memory is suitable for the template-matching method, which fetches pixel data many times, because the texture memory is a kind of cache memory.

The acquired frame image and the template image were copied into a texture memory on the GPU from the host computer. These images were divided into many block images with a size of 16 × 16, and each pixel in a block image was assigned to a thread of an SP. At the same time, a two-dimensional array that was used to store calculation results was copied onto the global memory of the GPU from the host computer, because the texture memory is a read-only memory.

To realize a template matching with the GPU, we assigned a thread to a pixel. In this way, the calculation of a block image was done with 16 MPs × 8 SPs, and 128 threads worked simultaneously. The CUDA works with a unit of 32 threads, and the maximum number of threads was defined as 512. The MSE calculation in each thread was processed in parallel, and the results were stored in a two-dimensional array on the global memory of the GPU. To improve the matching speed, we sampled only 1/16 of the pixels in the template image and calculated the MSE.

Lastly, the two-dimensional array on the global memory was copied into the host computer, and the position showing the best match on the temporal image was output with reference to the lowest MSE value in this array.
In our graphic user interface (GUI) that we developed, the extents of movement of three template images were displayed graphically as well as in three monitored images in acceptable time. The developed GUI is shown in Fig. 5. The extent of movement was calculated in mm converted from pixels, in which we quantified the relationship between an object-detector distance and the size of a pixel in an acquired image beforehand. The extent of movement was displayed on a monitor. In this system we treated the movement of the head such as the displacement in the horizontal and/or vertical direction, and the rotation of the head. These movements cause the displacement of the nose and ears on acquired images. In our preliminary study with a volunteer with the same situation as an actual treatment, the regular movement of a patient was expected to be less than 1.5 mm in all directions even though his head moves horizontally or vertically, or rotates slightly. The maximum acceptable limit of the displacement in the clinical situation is 2 mm, and so we targeted to detect an irregular movement beyond 2 mm. After placing a patient on the linac couch, we set the camera arm and measure the distance between the camera and the patient’s face. These data are input into the system and used for conversion of the distance from pixels to mm. Next, we acquire three initial images that are used for making template images. We set a rectangular region of interest (ROI) for each image manually and trim an image. These three images (nose and both ears) become the template images for three cameras. We start our monitoring system, and the stereotactic irradiation starts at the same time. If the patient’s head moves beyond 2 mm on at least one of three temporal images, the system alerts a radiologist with an alarm beep and shows the message to turn the beam off. Fig. 5 shows a screen shot of the GUI.

3. Experiments

1) Robustness to illuminance change

To validate the performance of our proposed system, we conducted some experiments under the same conditions as those used in daily stereotactic radiation therapy. Fig. 6 shows our radiation therapy room with a linac (CLINAC iX, Varian Medical Systems, California, USA). In the room there were two fluorescent lamps on the ceiling. The robustness of our motion detection system to the illuminance change of the targeted areas (the nose and both ears) was confirmed as follows: We first measured the illuminance of the targeted areas with an illuminance meter (MT-8210, MotherTool, Japan) while rotating the gantry of the linac. During the rotational movement of the linac gantry,
the shadow of the gantry appeared on a volunteer’s head. In this experiment, we rotated the gantry from 0 to 180 degrees in 10 degree-steps, and we measured the illuminance of the targeted areas. In addition, we measured the extent of the movement by averaging the results for 50 frames (5.9 seconds). The movement of the gantry is shown in Fig. 8. In addition, to validate the repeatability, we conducted a similar experiment while rotating the gantry inversely from 180 to 0 degrees. The size of an image was QVGA (320 × 240) in this experiment, and the distance between the camera and the volunteer’s nose or ears was 5 cm (1 pixel = 0.180 mm). To compare the results with those with visible light, we measured the movement of the same area without the IR filter and without infrared light.

2) Accuracy of measurement with a head phantom

To evaluate the accuracy of the detected movement, we conducted the experiment with a head phantom with a z-axis stage or a rotational table. First, we put the head phantom on a z-axis stage (TSD-1203, SIGMA KOKI, Tokyo, Japan) and set three USB cameras around the phantom as shown in Fig. 7. To imitate the clinical situation, each camera covered the nose and both ears. In this experiment, we moved the stage downward by 10 mm vertically at a speed of 1.0 mm/sec and then moved the stage upward by 10 mm at the same speed. Next, we put the same phantom on the rotational table (SAK80-R, Sakurai Manufacture Place, Saitama, Japan) and rotated the phantom from 0 to 3 degrees clockwise, and 6 degrees counterclockwise, and then 3 degrees clockwise. In this experiment we placed three markers on the face of the phantom and measured the true movement in the horizontal direction, because the phantom moved rotationally, precluding use of a rotation angle as an index of the horizontal movement of the phantom. The accuracy was evaluated with the horizontal movement of the targeted area.
3) Accuracy of measurement with a volunteer

We evaluated the detection accuracy of the movement with a volunteer by using the same geometry as a clinical setting without the linac gantry. In this experiment, we also used a bite plate, and we asked the volunteer to relax during the acquisition of images for 5 min. To measure the true movement of the volunteer’s head, we used three markers and put them on the volunteer’s face. We measured the extent of the movement by manual tracing for each frame, using a marker, and compared it with the output of the system that was calculated automatically. The input image size was VGA (640 × 480). The distance between the camera and the patient’s nose was 4.5 cm (1 pixel = 0.074 mm), the right ear 8 cm (1 pixel = 0.130 mm), and the left ear 7 cm (1 pixel = 0.116 mm).

4. Results

1) Robustness to illuminance change

**Fig. 8** shows the variation of the illuminance of the ears and nose. In (a)-(c), the gantry moved from 0 to 180 degrees. Depending on the gantry angle, the illuminance (right scale) changed from 20 to 270 lux. The error in the detected positions is also shown (left scale). In (d)-(f), the gantry moved from 180 to 0 degrees. In these graphs, the variation of the illuminance and the detection error are shown as in the above graphs. The error caused by the illuminance of the visible light was significant under a dark condition, becoming almost zero under the infrared lighting. **Table 1** shows the mean absolute error, standard deviation, and maximum absolute error in this experiment.

2) Accuracy of measurement with a head phantom

**Fig. 9** shows the accuracy of our motion detection system with use of the head phantom. In (a)-(c), we moved the head phantom with a constant speed in the direction of the z-axis. The size of the template image was 120 × 120 and the frame rate acquired was 4.07 fps. (a) is the result for the right ear, (b) for the nose, and (c) for the left ear. The

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Mean absolute error, standard deviation, and maximum absolute error of detected movements under the illuminance change in a radiation room (mean±SD[max] in mm).</th>
</tr>
</thead>
<tbody>
<tr>
<td>moving direction</td>
<td>from 0 to 180 deg.</td>
</tr>
<tr>
<td>lighting condition</td>
<td>visible light</td>
</tr>
<tr>
<td>right ear</td>
<td>9.5±12.6 [29.6]</td>
</tr>
<tr>
<td>nose</td>
<td>16.7±13.6 [35.0]</td>
</tr>
<tr>
<td>left ear</td>
<td>8.10±6.20 [15.7]</td>
</tr>
</tbody>
</table>

**Fig. 8** Illuminance change and detection error under visible lighting and infrared lighting. (a), (b), and (c) are results for rotating the gantry from 0 to 180 degrees. (d), (e), and (f) are results for rotating the gantry from 180 to 0 degrees.
Table 2 Mean absolute error, standard deviation, and maximum absolute error of detected movements of the phantom (mean±SD[max] in mm).

<table>
<thead>
<tr>
<th></th>
<th>Vertical movement (113 frames)</th>
<th>Rotational movement (97 frames)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>horizontal direction</td>
<td>vertical direction</td>
</tr>
<tr>
<td>right ear</td>
<td>0.133±0.145[0.475]</td>
<td>0.148±0.181[0.725]</td>
</tr>
<tr>
<td>nose</td>
<td>0.102±0.170[0.633]</td>
<td>0.132±0.148[0.623]</td>
</tr>
<tr>
<td>left ear</td>
<td>0.195±0.203[0.475]</td>
<td>0.135±0.149[0.678]</td>
</tr>
</tbody>
</table>

dotted line shows the movement of the head phantom. (d)-(f) are the results of the rotational movement. The size of the template image was 120 × 120 and the frame rate acquired was 4.13 fps. (d) is the result for the right ear, (e) for the nose, and (f) for the left ear. In (e), the error becomes large in the case of the nose image, but the extent was less than 0.9 mm. Table 2 shows the mean absolute error, standard deviation, and maximum absolute error in this measurement.

3) Accuracy of measurement with a volunteer

Fig. 10 shows the extent of the movement measured with our system and the extent of true movement measured manually with a marker. (a), (b), and (c) show the results for the right ear, nose, and left ear, respectively. The mean absolute error and standard deviation of the detection error were 0.108±0.080 mm (right ear), 0.049±0.049 mm (nose), and 0.076±0.057 mm (left ear) in the 984 frames. The maximum absolute errors were 0.364 mm, 0.279 mm, and 0.260 mm, respectively.
5. Discussion

We proposed a contactless motion detection system for stereotactic radiosurgery, in which we used infrared USB cameras and the template-matching method. We adopted the infrared USB cameras and infrared lighting to realize a system, that were robust to the illuminance change due to the linac gantry. As regards the variation of the illuminance of the head position, the measured minimum illuminance was 20 lux and the maximum one was 270 lux according to the geometric relationship between the linac gantry and fluorescent lamps. These variations affect the detection accuracy of movement if we use visible light.

In this experiment, we compared the detection accuracy of the measurement with and without infrared light. The results showed that our motion detection system with the infrared light achieved good results within average errors of 0.26 mm (ears) and 0.62 mm (nose), whereas the motion detection with the visible light was inaccurate within an average error of 10 mm (ears) and 18 mm (nose). Our proposed system with the infrared light was insensitive to any changes in illuminance, and it detected the target movement within an error of less than 1 mm with an image size of 640 × 480 (VGA). In clinical situations, the positioning error of an x-ray beam in an irradiation should be suppressed to within 2 mm; therefore our system is thought to be applicable to actual radiation therapy.

For applying our system clinically, the calculation time required for the detection of movement is also very important. Our system realized the template matching with a frame rate of about 7.62, that is, the time needed for
calculation was 0.131 sec for three frames. This means that the calculation time of a frame for each camera was 0.043 sec. By using OpenMP [16] and assigning the calculation tasks to three threads, we can achieve a faster calculation than that with a common CPU (Intel Core i7 920XM Extreme Edition, 2 GHz, maximum performance 32 GFLOPS, measured performance 25.3 GFLOPS). The performance of GPU (NVIDIA Quadro FX 3800M, 675 MHz) used in this system was 648 GFLOPS in the maximum performance and the performance ratio was 79.08 times compared to the CPU. However, the use of the GPU enabled us to perform 11.62 times faster calculation than that with the OpenMP, which is compatible with the calculation speed that was approximately 50 times faster than that with the CPU. We think that a frame rate of five is sufficient to avoid undesired irradiation with an x-ray beam due to sudden patient movement. From these points of view, our system is applicable for stereotactic radiation therapy with a sufficient response time.

6. Conclusion
We developed a contactless motion detection system with a GPGPU, and we evaluated the performance of the system under the same conditions as those used in actual radiation therapy. The results of the experiments showed that our system could detect the motion of the targeted objects within a mean absolute error of less than 0.7 mm under a change in illuminance, and the time required for calculation was 0.131 sec for three images. The results suggested the feasibility of its clinical application.

References
定位放射線治療は高エネルギーの細いX線ビームを用いて腫瘍を治療する非侵襲的な治療法である。この治療においては予期しない患者の動きは、近隣の正常組織や臓器に対して好ましくない結果となるため、我々は頭頸部の定位放射線治療のためのUSBカメラを用いた非接触動き検出システムを提案する。このシステムでは3台のUSBカメラによって得られた患者の鼻と両耳の3枚の動画像を用い、テンプレートマッチング法によって患者の動きを検出し、患者の動きが検出されたならばシステムが放射線技師や医師に対して、ビームをオフにするような警告を発する。我々は、対象領域を赤外光で照明し、同時にUSBカメラのレンズの前面でフィルタによって可視光領域の光を遮断した。動きの検出では汎用のGPUを用いテンプレートマッチング法を実装した。本論文ではこの動き検出システムの概要および有効性の検証結果を述べる。

キーワード：定位放射線治療、動き検出、テンプレートマッチング、赤外光、GPU、USBカメラ