# Accuracy of a stereo vision system for portable roadside portal monitoring *

Anonymous WACV submission

Paper ID 149

## Abstract

*Increasing concern over the smuggling of fissile material has prompted recent interest in constructing portable, rapidly deployable portal monitors to detect gamma radiation in passing vehicles. In this paper, we introduce a stereo system for precise, real-time, roadside tracking of vehicles in multi-lane roadways. The stereo system reports 3-D vehicle tracks to a gamma imager, which uses the track information to compensate for vehicle motion and produce detectable gamma peaks. We first describe the algorithmic components of the system, including calibration, camera motion correction, point tracking, stereo correspondence, and vehicle clustering. We also describe the physical system configuration. We then report results from tracking two different vehicles at speeds ranging from 5 to 20 MPH. We report on the accuracy of the system, where we observed a worst-case tracking error of 31 cm and an average tracking error of 9 cm in the direction of motion from a viewing distance of approximately 15 m.*

## 1. Introduction

Increasing concern over the smuggling of fissile material has prompted recent interest in constructing portable, rapidly deployable portal monitors to detect low-level gamma radiation in passing vehicles. Unlike traditional portal monitors, these systems would be deployed on the side of the road and would avoid impeding the flow of traffic; however, this operating condition presents a major challenge to successful gamma detection, since a single detector cannot simultaneously provide vehicle specificity and sufficient gamma efficiency to detect low-level sources at normal (e.g., interstate) traffic speeds. A coded-aperture gamma ray imager can provide a gamma "image" with high efficiency, but without compensating for vehicle motion, any peaks in the resulting gamma images would be heavily blurred and indistinguishable from background radiation [8].

Figure 1. Photograph of our two deployed instruments on either side of a roadway with fully extended 50-ft masts.

In order to overcome these challenges, we have designed and constructed a system (shown in Fig. 1) that couples gamma-ray and stereoscopic visible-light imaging in order to correct for vehicle motion in the gamma image and link every such image with a specific vehicle. In order to achieve successful motion correction, the vision system must track each passing vehicle in real time with high accuracy, and the resulting track must be transformed into the coordinate system of the gamma-ray imager. For this reason, we employ a stereo vision system using a camera pair mounted to a 50-foot mast extended vertically from a trailer on the ground. The height of the mast allows the cameras to see multiple lanes of traffic, and partially obscured vehicles can be differentiated by their 3-D position reported by the stereo algorithm. We use a Lucas-Kanade-based optical flow point tracker [2] to track features on each vehicle in one view and a correlation-based method to find corresponding points

*Notice: This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the U. S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

in the second view. In order to correct for mast motion, we employ a third camera mounted between the two stereo cameras that images corner-shaped fiducials affixed to the top of the trailer. We estimate the position of the mast in each frame by finding the position and orientation of the roof plane of the trailer relative to the third camera. Finally, we cluster the resulting stereo tracks into vehicles using a recently developed method known as "coherent motion regions" [3].

In this paper, we first describe our overall stereo vision system, including physical setup, calibration requirements, and algorithmic components. We then report results from tracking 12 vehicle passes from two different vehicles at speeds ranging from 5 to 20 MPH. Because gamma image quality depends largely on precision in the direction of vehicle motion, we measure the tracking accuracy of our system along this axis. We show that the average tracking error along the direction of motion is approximately 9 cm, which provides for sufficient motion compensation in the gamma imager. Note that the accuracy is based on gamma pixel size and is independent of velocity. Because this paper focuses on the visible-light imaging system, we do not show gamma-ray images.

## 2. Background

Many different methods have been proposed for separating moving foreground objects from background, and this is an active area of research with a large body of literature [5]. In our work we identified three main approaches: (1) background subtraction, (2) classification-type methods, and (3) feature detection and tracking. For our application we chose the latter approach due to the difficulty of learning an effective background model in a rapid-deployment system as well as the variability of vehicle appearances (especially from side views). In addition, methods for point detection and tracking are readily available as open source code and have been widely accepted in the computer vision community [1].

Initial work in this area culminated in the demonstration of a monocular system capable of detecting and locating the position of vehicles in fairly low-density regions, including crossing vehicles [8]. We introduce a more complex system in this paper, including imagers on both sides of the road, to defeat confounding factors of occlusion and improve detectability. The system described in this paper employs stereo vision sensors for three main reasons. First, the use of stereo vision enhances our ability to co-align two imagers on both sides of the road. Second, stereo vision should improve our general clustering and tracking ability by better identifying the range to vehicles, i.e., which lane each vehicle is in. Finally, the ability to detect the height of points on vehicles helps improve our ability to cluster and identify larger vehicles and possibly identify sources of occlusion.

## 3. Technical approach

We constructed two identical roadside instruments, each consisting of a coded-aperture gamma-ray imager and a stereo vision system mounted to a trailer. In this section, we first describe the physical setup of a single stereo vision system. We then describe the algorithm used to track vehicles in 3-D, including the optical flow, stereo matching, mast motion correction, and calibration procedures.

### 3.1. Physical configuration

Each stereo vision system is mounted on top of a retractable mast attached to the trailer that extends a maximum of 15.2 m in height. The stereo mount at the top of the mast is constructed from aluminum pipe and holds three cameras. Two cameras are mounted at either end of the mount separated by a distance of 2 m. These cameras, used for stereo vision, have a 95-degree diagonal field angle and are aimed at the road (assuming a height of 15.2 m), providing approximately an 24-m wide field of view in the closest lane of traffic. We refer to the left camera (looking from the trailer toward the road) of the stereo pair as the "reference" camera and the right camera as the "stereo" camera.

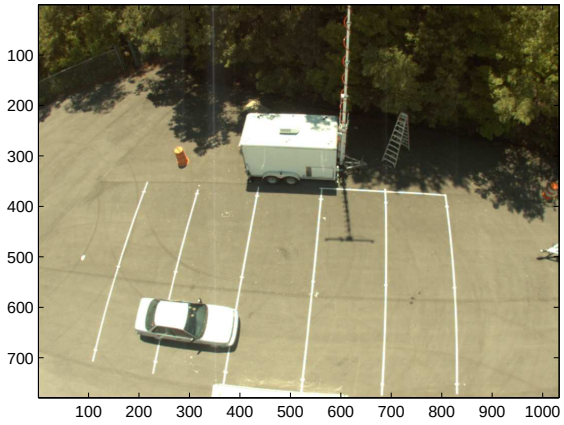
The third camera, used solely for mast-motion correction, is mounted close to the center of the mount and is aimed at the roof of the trailer. We refer to this camera as the "alignment" camera. In order to measure mast motion, we affixed several fiducials to the roof of the trailer, which are in the field of view of the third camera.

The cameras used in the system are Gevicam GP-series Gigabit Ethernet video cameras with $1032 \times 779$ resolution. We connect the three cameras to a single computer using 23-m CAT6 Ethernet cables. We were able to run the cameras continuously at 20 frames per second (FPS).

### 3.2. Stereo tracking algorithm

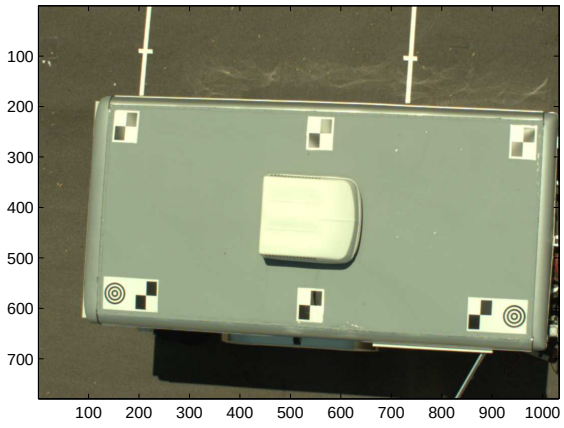
We first track feature points in 2-D using only images from the reference camera. We then use the corresponding stereo image to compute the 3-D location of each reference point. The result is a set of 3-D tracks, where some points in the track may be missing if a stereo correspondence could not be found. We describe each of these components in more detail below.

We use a standard Lucas-Kanade (L-K) optical flow-based point tracking algorithm to track feature points in the reference images in 2-D. For each frame, we first apply a corner detection algorithm based on minimum eigenvalue (as implemented in the OpenCV library) to find all the strong corner features in the image [7]. We then compute the optical flow vector at each corner point using a




(a)

(b)




(c)

Figure 2. Sample frames from each of the three cameras in the stereo vision system: (a) reference camera; (b) stereo camera; (c) alignment camera.

sparse pyramid-based version of the L-K optical flow algorithm [2], indicating the new location of the point in the next frame. We discard stationary points, i.e., those that moved less than a threshold distance between two consecutive frames. This process results in a set of 2-D point tracks, where each track ends when the point either leaves the image or becomes stationary.

For each point tracked in the reference image, we search for a corresponding point in the stereo image using normalized correlation. First, both the reference and stereo images are rectified so that all the epipolar lines are horizontal. Then for each reference image point, we search along the epipolar line in the stereo image by carrying out a normalized correlation in the $x$-direction. We need only correlate over the range of feasible disparities, which in our case can be computed using knowledge of the road boundaries. In order to account for small errors in the rectification, we correlate over a $3 \times N$ window, where $N$ is the number of pixels spanning the disparity range in the $x$ direction.

We select as the matching point the location of the correlation peak, i.e., the pixel with the highest correlation value. Two criteria are employed to determine the quality of match for each point correspondence. First, the value of the peak itself is compared with a threshold. Second, we compare the peak value to the second-highest correlation value in the window, excluding a $3 \times 3$ window around the peak, and we compute the following ratio:

$$\frac{c_1 - c_2}{c_1 - c_{min}} \tag{1}$$

where $c_1$, $c_2$, and $c_{min}$ are the highest, second-highest, and lowest correlation values in the window, respectively. We compare this ratio to a threshold. If either this or the previous criterion fail, we declare a non-match and hence do not compute the 3-D coordinates of the point.

Once the 3-D tracks have been computed, we cluster the tracks into vehicles using a method called "coherent motion regions" (CMR) [3]. CMR uses a sliding window to associate tracks whose position and motion are similar. The output of the algorithm is a set of cluster IDs for each frame, one ID for each track present in the frame. In order to accommodate streaming input, i.e., produce output prior to observing the full tracks, CMR processes frames in batches of a user-defined size.

### 3.3. Mast motion correction

The 15.2-m mast used to raise the stereo system is prone to flexing motions in windy or turbulent conditions, which, without software correction, can lead to errors in calculating absolute 3-D coordinates. In order to correct for this motion, we use the roof of the trailer as a reference plane and estimate its position relative to the alignment camera at every frame. Since the alignment camera is rigidly attached to the stereo mount and the roof plane remains approximately stationary with respect to the ground, the position of this reference plane can be used to compute the relative

motion of the stereo cameras with respect to the ground. More specifically, we ultimately compute a transformation $\mathbf{T}_{RF}$ that transforms points in the coordinate system of the reference camera to the roof plane coordinate system.

We use fiducial markings affixed to the trailer roof to estimate the roof plane. There are two circular target shapes and six corner shapes. In each alignment frame, we first locate the circular fiducials, whose appearance are invariant to any torsional rotation of the mast. We then use these as a reference to guess the approximate location of the six corner fiducials, whose coordinates relative to the circular fiducials are known. We use a corner-finding algorithm provided by OpenCV to refine the guess of each corner location to sub-pixel accuracy. Having located all six corners, we estimate the position and orientation of the roof plane relative to the alignment camera using an OpenCV library function. We use this to compute a transformation $\mathbf{T}_{AF}$ from alignment camera coordinates to roof plane coordinates. We multiply this by the transformation $\mathbf{T}_{RA}$ from reference to alignment obtained from calibration to yield $\mathbf{T}_{RF}$ as follows:

$$\mathbf{T}_{RF} = \mathbf{T}_{RA}\mathbf{T}_{AF}. \tag{2}$$

This transformation is updated every frame, so that any mast motion is automatically absorbed by the transformation from reference camera to roof plane.

### 3.4. Calibration

There are five main types of calibration that need to be carried out in our system: (1) intrinsic camera calibration, (2) camera-to-camera calibration, (3) camera-to-trailer calibration, (4) camera to gamma-ray imager calibration, and (5) trailer-to-trailer calibration. Intrinsic camera calibration is first performed for each camera individually. Multiple images are taken of a checkerboard calibration pattern at various poses with camera calibration software from DLR [6] used to calculate intrinsic camera parameters. These parameters include focal length, optical center, and lens distortion. After individual camera calibration, stereo calibration is next performed on the reference/stereo camera pair and the reference/alignment pair. Images are acquired with the checkerboard pattern simultaneously visible in each image pair. Checkerboard corner locations are extracted from the images using the DLR CalLab program, and then we perform an optimization to find the best fit camera-to-camera transformation that minimizes the projected RMS image error.

Camera-to-trailer calibration includes measuring the absolute coordinates of the corner-shaped fiducials on the trailer roof with respect to some user-defined origin. We also measure the positions of the corner-shaped fiducials relative to the circular fiducials in order to accurately guess the locations of the corners in each frame once the circular fiducials have been detected. A separate calibration is carried out to determine the coordinate system of the gamma-ray imager with respect to the trailer. This calibration is used to transform vehicle tracks into gamma-ray pixel coordinates for vehicle motion correction.

A final calibration step determines the transformation between the remote trailer reference frame and the local trailer frame. With the camera set for each trailer in position on the full raised mast, images are taken with each camera set and landmark points visible from both the remote and the local cameras are manually selected. The 3D coordinates of each point are calculated relative to the individual trailer. A method given by Horn [4] is then used to calculate the 3D transformation from the remote to local trailer coordinates.

## 4. Experimental setup

### 4.1. Physical setup

We set up two instruments to perform stereo tracking, one on either side of a two-lane roadway. The masts were raised to a height of approximately 15 m above the ground. Due to the east-to-west orientation of the roadway, one instrument was located to the north of the roadway looking south and the other was located south and looking north. This orientation caused the cameras on the north-side instrument to see more sun-induced glare effects. Images of the setup as seen from the top of the north-side mast are shown in Fig. 2.

Two different vehicles were driven along the roadway between the two instruments in the north lane at separate times: a 2006 Mazda 5 (Vehicle 1), and a 1992 Toyota Corolla (Vehicle 2). Each vehicle passed through the field of view six times, three eastward and three westward. Vehicle speed ranged from 5 MPH to 20 MPH. We acquired images from all three cameras at 20 FPS and applied our stereo tracking algorithm to the entire set of video data.

We placed six marker lines across the roadway perpendicular to the direction of travel at intervals of 10 ft, yielding a total of five 10-ft intervals. For each vehicle pass, we manually marked the frames, to half-frame accuracy, in which the vehicle's front tires crossed each line. This provided a certain amount of ground truth for the distance of travel for each vehicle pass.

### 4.2. Accuracy measurements

Because it is difficult to determine ground truth for absolute 3-D coordinates of objects with respect to a virtual coordinate system, we measure the accuracy of our system using relative 3-D positions, i.e., object displacement between two time instants. For each vehicle pass, the resulting 3-D tracks were clustered using the CMR algorithm, and only one of the resulting clusters corresponded to the

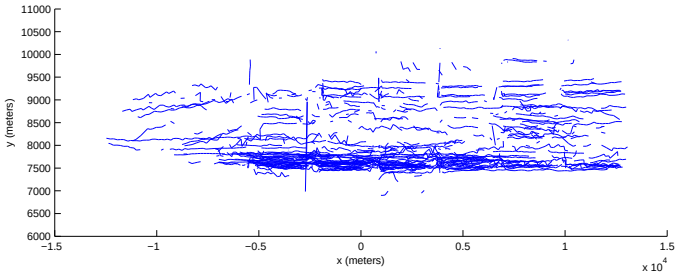



Figure 3. Overhead view ($x$-$y$ projection) of tracks for one pass of Vehicle 1.

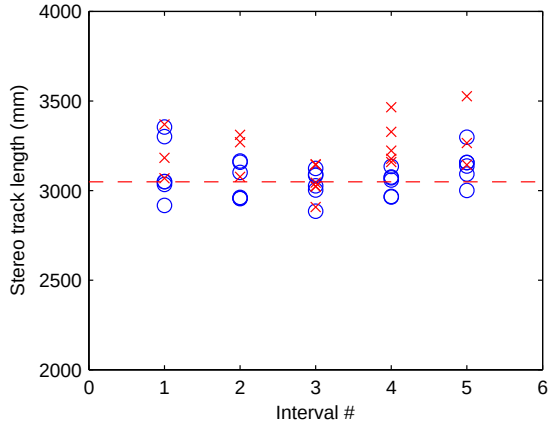
vehicle. All tracks not associated with this cluster were discarded. Tracks from one pass of Vehicle 1 corresponding to the vehicle cluster are shown in Fig. 3. For the remaining tracks, some of the points in the track were missing due to a poor match in the stereo correspondence algorithm.

For each of the five 10-ft intervals, we collected all the vehicle tracks which completely spanned the interval with no missing points (we refer to these as "valid" tracks). We then computed a single track equal to the spatial mean of all such spanning tracks. For the resulting mean track, we selected the two points from the track corresponding to the times at which the vehicle crossed the beginning and ending boundaries of that interval. The portion of the track between these two points is referred to as a "subtrack". We computed the Euclidean distance between the endpoints of the subtrack, and we refer to this as the "length" of the subtrack. This length ideally should be equal to 10 ft, assuming that the vehicle traveled a straight path perpendicular to the marker lines. Thus, for each interval and for each vehicle pass over that interval, we obtained a distance measurement of that interval from our tracking system that can be used to assess the accuracy of the system.

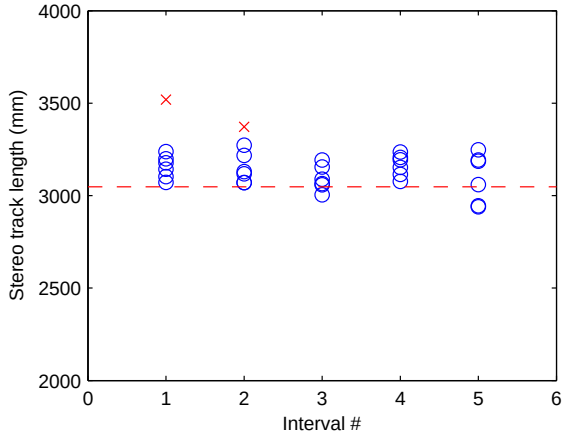
## 5. Results

Statistics of the valid tracks from both vehicles are shown in Table 1, and their lengths are plotted in Fig. 4. The table shows the mean and standard deviation of the subtrack lengths, the average error of each subtrack (i.e., the absolute difference between 10 ft and the length of the subtrack), and the number of valid subtracks obtained. For the South instrument, our system achieved a worst-case tracking error of 31 cm and an average tracking error of 9 cm in the direction of motion from a viewing distance of approximately 15 m.

The North instrument performed significantly worse, as evidenced by the low number of valid tracks obtained and the average track error. We observed that this performance difference was mainly due to sun glare effects, because the angle of the sun was such that reflections off of the vehicle surfaces were much more prominent in the North camera views. Bright reflections on the vehicle (see Fig. 5), were often tracked by the L-K point tracker, and these points typically caused highly inaccurate stereo correspondence matches due to the difference in location of the reflection in the two stereo views.

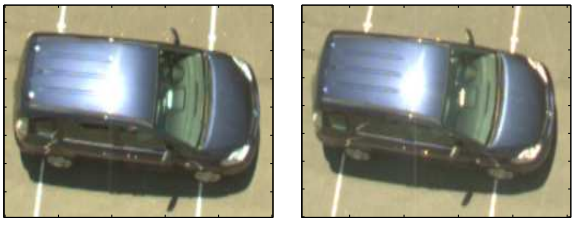



(a)




(b)

Figure 4. Scatter plots of track length for each of the five 10-foot intervals traversed by (a) Vehicle 1 and (b) Vehicle 2. Tracks from the north and south side instruments are denoted by 'x' and 'o', respectively, and the 10-ft (3048mm) ground truth point is shown as a dashed line.

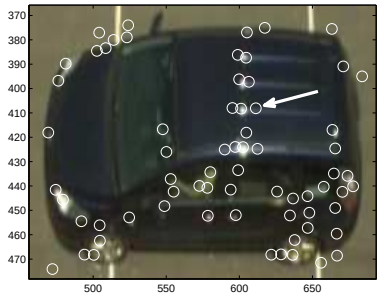


(a) (b)

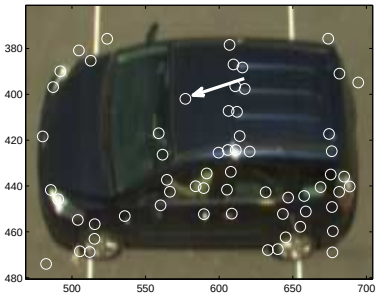
Figure 5. Example of inconsistent sun reflection pattern between the (a) left and (b) right stereo views in a particular frame that causes inaccurate stereo correspondences. Note how the position of the reflection differs relative to the physcial features on the roof.

Table 1. Subtrack statistics

| | | Subtrack length $\mu$ (mm) | Subtrack length $\sigma$ (mm) | Avg. error (mm) | # valid subtracks |
|---|---|---|---|---|---|
| Veh. 1 | North | 3193 | 209 | 144 | 64 |
| | South | 3077 | 113 | 29.5 | 521 |
| Veh. 2 | North | 3446 | 404 | 397 | 2 |
| | South | 3131 | 117 | 82 | 858 |




(a)




(b)

Figure 6. An extreme example of point tracking drift. The feature point called out by the white arrow in the upper frame (a) jumps far to the left in the next frame (b).

Other algorithmic factors that contributed to decreased tracking accuracy included point drift in the 2-D tracking algorithm and tracked points that are not strictly on the vehicle. Fig. 6 shows an extreme example of point drift in the L-K optical flow tracker that lead to tracking inaccuracies. This figure also shows examples of features on the road just above the vehicle that were tracked. Such points, while identified as corners by the minimum eigenvalue criterion, often lack the detail needed for a precise match in the stereo correspondence step, leading to significant errors in the calculated 3-D coordinates. These non-vehicle points also cause track discontinuities when the vehicle passes over the white marker lines, as can be seen in Fig. 3.

Small inaccuracies in the ground truth also contribute to error. When the vehicle crossed the outer marker lines near the sides of the image, its non-profile orientation with respect to the camera led to difficulty in judging the exact frame at which the front tires crossed the line. At the higher vehicle speeds (e.g., 20 MPH), an error of only one frame corresponds to more than 30 cm of distance.

## 6. Conclusions

Although we do not show gamma results in this paper, the accuracy of our stereo tracking system in these experiments is sufficient to obtain sharp, detectable gamma image peaks via motion correction for low-level sources. There are, however, many potential sources of error, including calibration, tracking, stereo correspondence, and mast motion; improvement in any of these areas would lead to a better tracking system and hence improved radiation detection. It is not clear from our experiments which individual error sources contributed the most to the overall error. In particular, we did not measure the effects of mast movement separately nor the effectiveness of our mast motion correction algorithm. Future experiments should measure all these factors independently. Also, higher vehicle speeds should be tested in order to more accurately reflect typical traffic speeds on major roadways.

## References

[1] OpenCV library. http://sourceforge.net/projects/opencvlibrary/, 2009. 2

[2] J.-Y. Bouguet. Pyramidal implementation of the lucas kanade feature tracer: Description of the algorithm. Technical report, Intel Research Laboratory, 1999. 1, 3

[3] A. M. Cheriyadat, B. L. Bhaduri, and R. J. Radke. Detecting multiple moving objects in crowded environments with coherent motion regions. In *Proceedings of the Sixth IEEE Computer Society Workshop on Perceptual Organization in Computer Vision (POCV)*, 2008. 2, 3

[4] B. Horn. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America*, 4:629–642, 1987. 4

[5] V. Kastrinaki, M. Zervakis, and K. Kalaitzakis. A survey of video processing techniques for traffic applications. *Image and Vision Computing*, 21(4):359–381, 2003. 2

[6] W. Sepp and S. Fuchs. DRL Camera Calibration Toolbox. 4

[7] J. Shi and C. Tomasi. Good features to track. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn.*, pages 593–600, 1994. 2

[8] K. P. Ziock, J. Collins, M. Cunningham, L. Fabris, T. Gee, J. Goddard, F. Habte, and T. Karnowski. The use of gamma-ray imaging to improve portal monitor performance. *IEEE Trans. Nuclear Science*, 55(6):3654–3664, 2008. 1, 2