Proceedings of the 29th Annual International
Conference of the IEEE EMBS
Cité Internationale, Lyon, France
August 23-26, 2007.

Monitoring Patient Respiration using a Single Optical Camera

Stefan Wiesner and Ziv Yaniv

Abstract— **We present a system for monitoring patient respiration using a single low cost camera. The system is based on tracking the motion of color fiducials placed on the patients abdomen. We show that our system is able to deal with variations in camera pose, and that its performance is equivalent to that obtained when using a commercial tracking system. We also validate our system using ultrasound images showing that it is sufficiently accurate for gating purposes in radiation treatment or for acquiring 4D (3D over time) data sets.**

I. Introduction

Monitoring patient respiration has proven to be a useful tool for dealing with the motion of anatomical structures in the regions of the abdomen and thorax in a variety of tasks. It has lead to improvements both in diagnosis, planning, and execution of interventions. In diagnosis, respiration monitoring has been used to improve the quality of PET, SPECT, and MR imaging by mitigating the effect of respiratory motion artifacts [1], [2], [3]. In radiation treatment planning, respiration monitoring has enabled improved dose distribution planning. By acquiring 4D CT data, 3D over time, plans can account for tumor motion throughout the respiratory cycle [4]. For interventional purposes, respiration monitoring in radiation therapy enables gated treatments, resulting in effective irradiation of tumor tissue while sparing surrounding healthy tissue [5], [6]. Finally, in CT fluoroscopy based needle biopsy procedures monitoring patient respiration has been used to provide quantitative breathing instructions to the patient, resulting in improved accuracy of needle placement [7].

Many devices are in use for monitoring patient respiration, with no specific one gaining wider acceptance than others. We categorize respiratory monitoring devices into two categories, devices that are invariant to rigid body motion and those that are not. For most tasks this invariance is not desirable, as respiration monitoring is used as a means to deal with patient motion. Devices that are invariant to rigid body motion make the limiting assumption that all motion is due to respiration, and should thus be used with caution.

Common respiratory monitoring devices that are inherently invariant to rigid body motion include devices based on the temperature of airflow (e.g. [1]), on direct measurement of airflow (e.g. [8]), and on the tension of a belt placed on the patient (e.g. [9]).

S. Wiesner is with the Imaging Science and Information Systems (ISIS) Center, Dept. of Radiology, Georgetown University Medical Center, Washington, DC, USA, and Chair for Computer Aided Medical Procedures (CAMP), TU Munich, Munich, Germany wiesner@isis.georgetown.edu

Z. Yaniv is with the Imaging Science and Information Systems (ISIS) Center, Dept. of Radiology, Georgetown University Medical Center, Washington, DC, USA zivy@isis.georgetown.edu

All devices that are not invariant to rigid body motion are based on monitoring the patient's chest motion using one or more cameras. These include the use of a monocular infrared camera system as described in [10], [4], the use of a stereo system using standard video images acquired by webcams as described in [11], and the use of an infrared stereo system as described in [12].

We propose to use standard video images acquired with a single webcam to perform respiratory monitoring. This solution is not invariant to rigid body motion, it is more affordable than the monocular infrared camera based approach, and it provides a simpler solution than the use of video based stereo reconstruction as it does not require any camera calibration.

II. Materials and Methods

Our respiratory monitoring system is based on automatic detection and tracking of the motion of color fiducials placed on the patient's abdomen in the epigastric region (near the xiphoid process) and the left and right hypochondriac regions. The system consists of a single webcam mounted on an adjustable arm and colored cube fiducials attached to a velcro belt, as shown in Figure 1. Each of the fiducial cubes faces has a different color, allowing the system to choose an appropriate color based on the colors present in the specific environment. Video images are acquired with a resolution of 640×480 at a frame rate of $30Hz$.

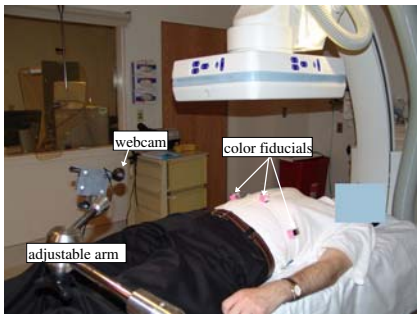

Fig. 1. Hardware components for respiratory monitoring: of the shelf webcam (Logitech, QuickCam Fusion), plastic cubes with colored faces, and adjustable arm.

The respiratory signal is generated from the tracking data of the fiducial with the largest motion magnitude along the principle axis of variance as defined by the Principle Component Analysis of the 2D motion. In practice this has always corresponded to the fiducial closest to the xiphoid process. The remaining fiducials are used in conjunction with

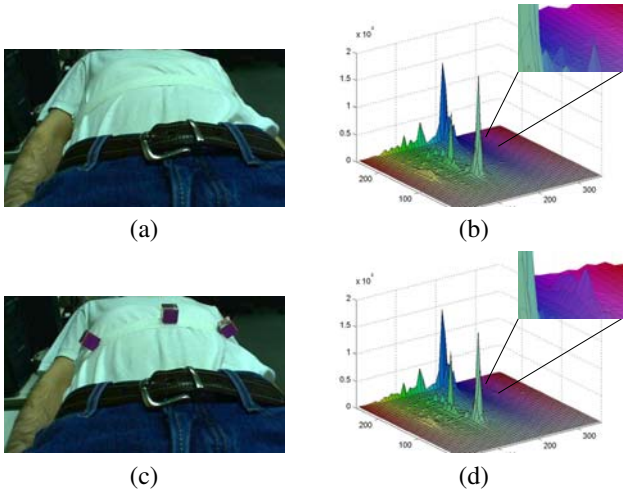

Fig. 2. Automatic selection of marker colors according to image's Hue-Saturation histogram: (a) image without marker (b) corresponding Hue-Saturation histogram with sparsest region corresponding to the purple Hue value, inset shows zoomed in region (c) image with markers (d) new Hue-Saturation histogram exhibiting a peak in the region corresponding to the purple markers, inset shows zoomed in region.

1) Initialize:
   a) Select fiducial color based on analysis of Hue-Saturation histogram of patient setup image.
   b) Place fiducial markers.
   c) Detect color blobs corresponding to marker color using thresholds on the Hue and Saturation values and perform connected component analysis.
   d) Initialize the CAMSHIFT algorithm using the detected blobs and track them for $30sec.$
   e) Perform PCA on the tracked data and project each blob's motion onto its principle axis.
   f) Analyze the frequency of the signal generated by each of the blobs, if it is in the range of 12 to 25 cycles per minute it is classified as a fiducial.
   g) Identify the fiducial with the maximal magnitude along its principle axis, the primary fiducial (PF).
2) Respiratory signal acquisition:
   a) Track blobs identified as fiducials in the previous stage using the CAMSHIFT algorithm.
   b) Project the location of the PF onto its principle axis as computed during initialization (this is the respiratory signal).

TABLE I

COLOR BASED MONOCULAR RESPIRATORY MOTION MONITORING FRAMEWORK.

this fiducial to improve the systems sensitivity to rigid body motion.

To robustly use color as the feature of interest we work in the Hue Saturation Value color space. This decouples the object's color from the intensity information, and when using the hue information provides relative invariance to illumination and viewing direction [13]. Thus the system does not require a strict relationship between the camera and patient positions, a useful characteristic in the clinical setting.

It should be noted that our system requires initialization. This is an automatic process during which we assume the patient is exhibiting normal tidal breathing. We now describe our approach in detail.

A. System description

As a first step we choose the best possible color for tracking in the specific environment from among the four colors on the cube's faces. An image of the patient setup is acquired, and the system indicates which of the four possible colors should be used. This is based on an analysis of the Hue-Saturation histogram, with the color of choice having the hue whose surrounding region in the histogram is the sparsest. From our experiments we have empirically determined that a region of $\pm 15^o$ is sufficient. Figure 2 illustrates this process.

Once the color is chosen, the fiducials are attached to the velcro belt and an image is acquired. Possible fiducial locations in the image are detected using a threshold on the the hue and saturation values followed by connected component analysis. The threshold on the hue values is set to the fiducial hue $\pm 15^o$ as was the case in the previous step. In addition we require a minimal saturation value of 0.3, as at low saturation values the hue data is less reliable for color based segmentation. Connected components in the resulting binary image are detected and those that consist of less than 50 pixels are considered noise, as marker sizes in the images vary between 400 to 900 pixels.

The locations of the centroids of the remaining connected components are candidate fiducial locations in the image. These serve as the initial input to the CAMSHIFT blob tracking algorithm [14]. The CAMSHIFT algorithm tracks the color blobs for 30 seconds and this data is used to detect the actual fiducial locations in the image.

We perform Principle Component Analysis (PCA) on the path of each tracked blob. The 2D points are projected onto the principle axis, the eigen-vector corresponding to the largest eigen-value, resulting in a one dimensional signal. For fiducials we expect this signal to be correlated with the respiratory cycle, that is, it will have a dominant frequency in the range of 12 to 25 cycles per minute, as respiratory rates outside this range are considered abnormal. All candidate fiducials that pass this test are assumed to be true fiducials.

We now identify the fiduicial that has the maximal motion along its principle axis. This fiducial is our Primary Fiduical (PF) and is used to generate the respiratory signal.

Once the initialization phase is completed the CAMSHIFT algorithm is used to track the fiduical locations in the image and a respiratory signal is generated by projecting the motion of the PF onto its principle axis which was computed during the initialization phase. Table I summarizes our approach.

III. EXPERIMENTAL RESULTS

We test the accuracy of our system by comparing the respiratory signal generated using the video images to a respiratory signal generated using the Polaris Vicra tracking system from Northern Digital Inc. (Waterloo, Ontario, Canada). This tracking system has a manufacturer specified spatial localization RMS error of $0.25mm$.

To generate both respiratory signals we attach markers tracked by the Polaris system next to our fiducials and simultaneously acquire video images and 3D tracking data. The respiratory signal based on the polaris data is generated by performing PCA on these 3D points, with the respiratory signal corresponding to the projection of the data onto the computed principle axis, the eigen-vector corresponding to the largest eigen-value. We then compare the two signals using normalized cross correlation, Pearson's r, as our similarity measure.

To evaluate our system's performance we acquired four types of data sets from a volunteer in which he exhibited tidal breathing, tidal breathing interrupted by coughing, tidal breathing interrupted by periods of breath-holding, and tidal breathing interrupted by motion. For each category we acquired three data sets. Figure 3 shows respiratory signals from each type of data set. As can be seen from this figure and the results presented in Table II the respiratory signal generated by the Polaris system and our video based system are highly correlated, with all correlation values above 0.95 except for whole body motion.

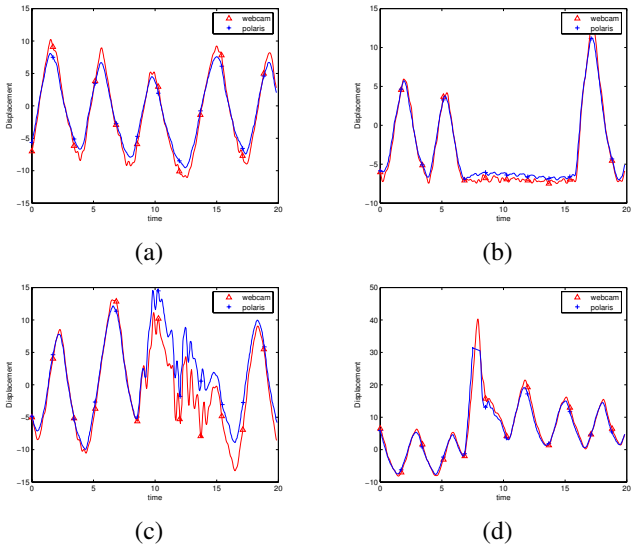

Fig. 3. Respiratory signal obtained by our method and using the Polaris Vicra for (a) normal breathing, NCC=0.99; (b) breath hold, NCC=0.99; (c) coughing, NCC=0.95; and (d) patient motion during acquisition, NCC=0.98. Several sample points are marked on each graph.

As we are using a single video camera to perform motion tracking we evaluate the effect of various spatial relationships between camera and patient on tracking, as these depend upon the clinical application. We acquired several data sets with varying camera locations. For each data set we generate a respiratory signal based on the Polaris data and video images. In all experiments the acquired data is of a volunteer exhibiting normal tidal breathing. The correlation between the Polaris based signal and video based signal was above 0.99 for all experiments. That is, the effect of varying camera pose is negligible. Figure 4 shows three examples.

Finally, we evaluated our system using a portable ultrasound (US) system, the Terason 2000 from Teratech Corp.

normal breathing	breath hold	coughing	patient motion
0.99	0.99	0.97	0.98
0.99	0.99	0.95	0.91
0.99	0.99	0.95	0.87

TABLE II

NCC VALUES FOR EVALUATING VIDEO BASED RESPIRATORY MOTION MONITORING VS. THE POLARIS INFRARED STEREO BASED SYSTEM. FOR EACH RESPIRATION CATEGORY WE ACQUIRED THREE DATA SETS. THE RESULTS SHOW A HIGH CORRELATION BETWEEN THE TWO APPROACHES, FOR ALL SCENARIOS.

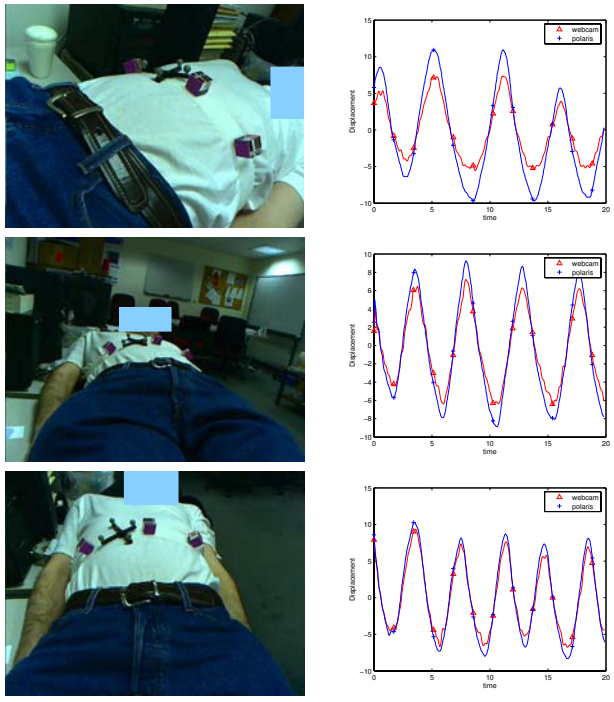

Fig. 4. Effect of camera placement on system performance, (a) sample image from video stream; and (b) respiratory signal obtained by our method and using the Polaris Vicra for a volunteer exhibiting normal tidal breathing. In all cases the NCC was above 0.99.

(Burlington, MA, USA). Figure 5 shows this setup. A volunteer was asked to perform normal tidal breathing and we generated a respiratory signal which was displayed on screen. We then arbitrarily chose a point in the respiratory cycle and asked the volunteer to hold his breath at that point in the cycle. The volunteer then returned to normal tidal breathing and was asked to hold his breath again when he reached the same point in the respiratory cycle as indicated by our display. This process was repeated multiple times (four or more per experiment). At each of the breatholds we acquired an US image. We then acquired US images at random points in the respiratory cycle using the same breathold technique. By comparing the NCC value for the images we show that our respiratory signal is consistent with organ movement due to respiration.

For all images acquired at the same point in the respiratory

cycle the correlation coefficient between all possible pairs was between 0.85 and 0.93. For all the images acquired at random points in the respiratory cycle the correlation coefficient varied between 0.44 and 0.68. Figure 6 shows sample US images from one of the experiments. To estimate the maximal possible correlation between two US images we acquired six images of a volunteers ulna without any motion between probe and anatomy. This was repeated for both hands. For all the possible image pairs the correlation coefficient was between 0.94 and 0.98, although in practice it should have been 1. That is with this US machine a perfect correlation for exactly the same underlying physical situation is unlikely.

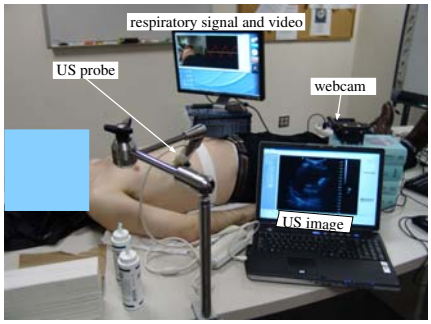

Fig. 5. Experimental setup for US based assessment of respiratory signal generation.

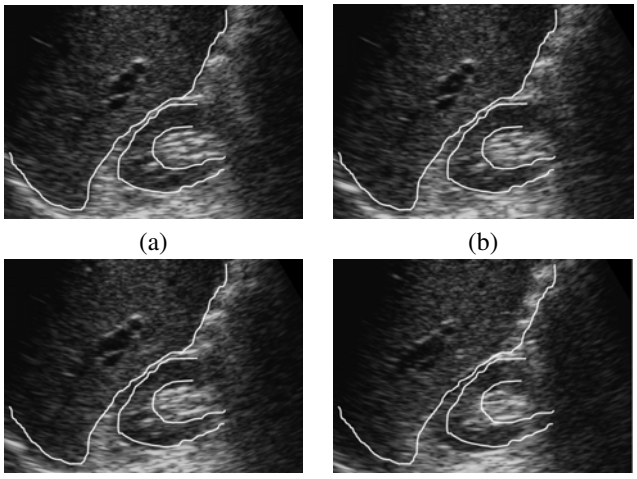

Fig. 6. US images acquired for validating the respiratory signal. Distinct structures are marked in image (a) which was acquired at a specific point in the respiratory cycle and transferred to the other images. Images (b), and (c) were acquired subsequently at that same point in the respiratory cycle and image (d) at a random point. It is clear that images acquired at the same point in the respiratory cycle as defined by our system are correlated.

IV. Discussion and Conclusions

We have developed a low cost system for monitoring respiration. The performance of our approach has been evaluated using a comparison study with a commercial infrared stereo system, and using US images for validating the usefulness of our system for gated interventions and image acquisition.

We have shown that the differences between our webcam based approach and using a commercial tracking system are negligible and that the respiratory signals generated by both methods are highly correlated, with a correlation coefficient above 0.95 for all respiration experiments, except in the case of patient motion.

Finally, we have shown that acquiring US images at the same point in the respiratory cycle as defined by our respiratory signal results in highly correlated images, with a correlation coefficient above 0.85 for all image pairs. This validates the usefulness of our approach to respiratory signal generation for gating purposes.

Acknowledgment

We would like to thank Prof. Nassir Navab, Prof. Kevin Cleary, and Dr. Filip Banovac for their help facilitating this research. This work was funded by US Army grants DAMD17-99-1-9022 and W81XWH-04-1-0078.

References

[1] L. Boucher, S. Rodrigue, R. Lecomte, and F. Bénard, "Respiratory gating for 3-dimensional PET of the thorax: Feasibility and initial results," *Journal of Nuclear Medicine*, vol. 45, no. 2, pp. 214–219, 2004.

[2] R. L. Ehman, M. . McNamara, M. Pallack, H. Hricak, and C. B. Higgins, "Magnetic resonance imaging with respiratory gating: techniques and advantages," *American Journal of Roentgenology*, vol. 143, no. 6, pp. 1175–1182, 1984.

[3] Y. Kawakami, K. Suga, T. Yamashita, H. Iwanaga, M. Zaki, and N. Matsunaga, "Initial application of respiratory-gated ^{201}Tl SPECT in pulmonary malignant tumours," *Nuclear Medicine Communications*, vol. 26, no. 4, pp. 303–313, 2005.

[4] T. Pan, T.-Y. Lee, E. Rietzel, and G. T. Y. Chen, "4D-CT imaging of a volume influenced by respiratory motion on multi-slice CT," *Med. Phys.*, vol. 31, no. 2, pp. 333–340, 2004.

[5] M. Chang, P. Lin, S. Benedict, H. Thames, S. Vedam, and P. Keall, "Investigating the effects of respiratory gated IMRT treatment delivery on survival," *International Journal of Radiation Oncology Biology Physics*, vol. 63, no. Supplement 1, pp. S386–S387, 2005.

[6] E. Yorke, G. Mageras, T. LoSasso, H. Mostafavi, and C. Ling, "Respiratory gating of sliding window IMRT," in *Engineering in Medicine and Biology Society*, 2000.

[7] S. K. Carlson *et al.*, "CT fluoroscopy-guided biopsy of the lung or upper abdomen with a breath-hold monitoring and feedback system: A prospective randomized controlled clinical trial," *Radiology*, vol. 237, no. 2, pp. 701–708, 2005.

[8] W. Kalender *et al.*, "Measurement of pulmonary parenchymal attenuation: use of spirometric gating with quantitative CT," *Radiology*, vol. 175, no. 1, pp. 265–268, 1990.

[9] S. K. Carlson *et al.*, "Intermittent-mode CT fluoroscopyguided biopsy of the lung or upper abdomen with breath-hold monitoring and feedback: System development and feasibility," *Radiology*, vol. 229, no. 3, pp. 906–912, 2003.

[10] S. A. Nehmeh *et al.*, "Quantitation of respiratory motion during 4D-PET/CT acquisition," *Med. Phys.*, vol. 31, no. 6, pp. 1333–1338, 2004.

[11] M. A. Gennert, J. K. Ho, A. C. Quina, J. H. Wang, P. P. Bruyant, and M. A. King, "Feasibility of tracking patient respiration during cardiac SPECT imaging using stereo optical cameras," in *Nuclear Science Symposium*, vol. 5, 2003, pp. 3170–3172.

[12] R. D. Beach, H. Depold, G. Boening, P. P. Bruyant, B. Feng, H. C. Gifford, M. A. Gennert, S. Nadella, and M. A. King, "An adaptive approach to decomposing patient-motion tracking data acquired during cardiac SPECT imaging," *IEEE Trans. Nucl. Sci.*, vol. 54, no. 1, pp. 130–139, 2007.

[13] T. Gevers and A. W. M. Smeulders, "Color based object recognition," in *International Conference on Image Analysis and Processing*, 1997, pp. 319–326.

[14] G. R. Bradski, "Computer vision face tracking for use in a perceptual user interface," *Intel Technology Journal*, no. Q2, pp. 1–15, 1998.