The HKU Scholars Hub   The University of Hong Kong   香港大學學術庫

Title	**Tracking Lumbar Vertebrae in Digital Videofluoroscopic Video Automatically**
Author(s)	**Wong, SF; Wong, KYK; Kris Wong, WN; John Leong, CY; Keith Luk, DK**
Citation	**The International Workshop on Medical Imaging and Augmented Reality (MIAR 2004), Beijing, China, 19-20 August 2004. In Lecture Notes in Computer Science, 2004, v. 3150, p. 154-162**
Issued Date	**2004**
URL	**http://hdl.handle.net/10722/103501**
Rights	**Creative Commons: Attribution 3.0 Hong Kong License**

Tracking Lumbar Vertebrae in Digital Videofluoroscopic Video Automatically

Shu-Fai WONG[1], Kwan-Yee Kenneth WONG[1], W. -N. Kris WONG[2]
, C. -Y. John LEONG[2], and D. -K. Keith LUK[2]

[1] Department of Computer Science and Information Systems, The University of
Hong Kong, Hong Kong, {`sfwong, kykwong`}`@csis.hku.hk`
[2] Department of Orthopaedic Surgery, The University of Hong Kong, Hong Kong

Abstract. Low back pain becomes one of the significant problem in
the industrialized world. Efficient and effective spinal motion analysis is
required to understand low back pain and to aid the diagnosis. Videofluo-
roscopy provides a cost effective way for such analysis. However, common
approaches are tedious and time consuming due to the low quality of the
images. Physicians have to extract the vertebrae manually in most cases
and thus continuous motion analysis is hardly achieved. In this paper,
we propose a system which can perform automatic vertebrae segmenta-
tion and tracking. Operators need to define exact location of landmarks
in the first frame only. The proposed system will continuously learn the
texture pattern along the edge and the dynamics of the vertebrae in the
remaining frames. The system can estimate the location of the vertebrae
based on the learnt texture and dynamics throughout the sequence. Ex-
perimental results show that the proposed system can segment vertebrae
from videofluoroscopic images automatically and accurately.

Key words: Motion Tracking, Spinal Motion Analysis

1   Introduction

Low back pain is one of the most common health disorders and its cost is enor-
mous [1]. There is a general consensus that the diagnosis and the treatment of low
back pain can be aided by analysing spinal movement [2]. Thus, spinal measure-
ment techniques have been studied widely. At present, videofluoroscopic imaging
provides an effective method of obtaining images for spinal motion analysis. Gen-
erally, landmarks of a moving vertebra will be extracted from videofluoroscopic
video and will then be analysed. Landmarks are usually the corners of the mov-
ing vertebra and are usually extracted manually [3]. Unfortunately, the analysis
is difficult and time consuming due to the low quality of the videofluoroscopic
images. Figure 1(a) shows typical videofluoroscopic image of spine. Thus, a wide
range of researches on automatic extraction of landmarks have been conducted,
such as [4].

   In general, there are two main approaches in videofluoroscopic analysis. The
first one is based on template matching and correlation (e.g. [5]) which is simple

to implement and easy to understand. However, such approach involves pixel-to-pixel comparison and thus susceptible to changing contrast and pixel intensity of the image. Another approach which is based on feature detection is adopted in current research. Features can be corners, edges and shape. In [6], the vertebrae in the images are located by matching corners. In [7], active shape models is used to improve the robustness by introducing shape constraints. To reduce the searching size, generalized hough transform is used in [8]. Such approach is computational efficient but making unrealistic assumption of high image contrast. Edges and features have to be manually enhanced and refined before feature location can be done. It seems that most of the commonly adopted approaches can be considered as computer-aided but not automatic.

In this paper, we propose a method in which an active contour (or snake) will attach to the vertebrae automatically throughout the video sequence. Users only need to define landmark positions on the first videofluoroscopic image. The active contour formed from such landmarks will attach to the vertebra automatically in the remaining video images. This greatly reduces the effort of physicians in setting accurate landmarks of vertebra manually in every video frame. The reduction in human intervention means the reduction in error rate due to fatigue of the operator. Analysis on spinal motion can be done much more effectively and accurately.

2   System Architecture

The whole system consists of three major modules, namely feature learning module, feature detection module and tracking module. The workflow of the system is shown in figure 1(b).

Given the first image and the exact position of the landmarks, the feature learning module will learn the texture pattern encoded by Markov Random Field (MRF) [9] using Support Vector Machine (SVM) [10] along the edge. On the other hand, a snake or active contour [11] is formed from the landmarks. When second image is input, the feature detection module will detect the edge using the texture information from feature learning module along the snake. The snake will then fitted toward features (or edges) detected. The tracking module will then learn the dynamic of those landmarks using Kalman filter [12]. It will predict the location of the landmarks in the next frame and shift the snake accordingly. At the same time, the feature learning module will learn the texture pattern again. The feature detection will then detect the features in the next frame as described above and the whole process repeats. Thus, given the videofluoroscopic video, the corresponding time series data on the location of the landmarks will be obtained by the system.

3   Implementation details

As described in previous section, there are several components within each module in order to achieve the aim of automatic vertebrae tracking. These compo-
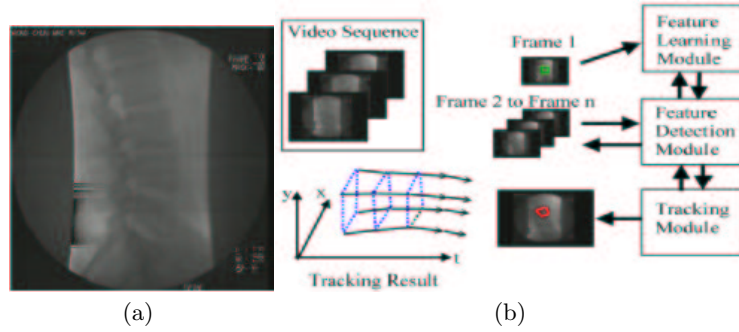
(a)   (b)

Fig. 1. *In (a), it shows a typical videofluoroscopic image of spine. In (b), it shows the workflow of the system.*

nents are shared among the modules and work collaborately. These components include the MRF texture descriptor, the SVM pattern recognizer, the snake and the kalman filter. The implementation details of these components will be explored in this section.

3.1 Texture description by Markov Random Field

Markov Random Field was first developed for texture analysis, e.g. [13]. It can be used to describe a texture and make prediction on the intensity value of a certain pixel given the intensity value of its neighborhood. The theories related to Markov Random Field can be found in [9].

In Markov Random Field, the neighborhood is defined as clique elements. Consider that $S = \{s_1, s_2, .., s_P\}$ is a set of pixels inside the image, and $N = \{N_s | s \in S\}$ is the neighborhoods of the set of pixels. In the system, the neighborhoods are the 8 pixels that with chessboard distance 1 away from the target pixel.

Assuming $X = \{x_s | s \in S\}$ is the random variables (the intensity value) for every pixel inside an image, where $x_s \in L$ and $L = \{0, 1, .., 255\}$. Besides, we have a class set for texture pattern, $\Omega = \{\omega_{S1}, \omega_{S2}, ..., \omega_{SP}\}$ where $\omega_{Si} \in M$ and M is the set of available classes. In the proposed system, we have only two classes, the edge and the non-edge classes.

In Markov chain analysis, the conditional probability of certain pixel being certain class is given by Gibbs distribution according to Hammersley-Clifford theorem. The density function is $\pi(\omega) = \frac{1}{\sum_\omega exp(\frac{-U(\omega)}{T})} exp(\frac{-U(\omega)}{T})$, where T is the temperature constant, which is used in stimulated annealing. The energy term can be further represented as $U(\omega, x_i) = V_1(\omega, x_i) + \sum_{i' \in N_i} \beta_{i,i'} \delta(x_i, x_{i'})$, where $V_1(\omega, x_i)$ represents the potential for pixel with certain intensity value belongs to certain class and the $\delta(x_i, x_{i'})$ is the normalised correlation between pixel at s_i and those at $s_{i'}$.

When the texture is being learnt by the feature learning module, the set of $\beta_{i,i'}$ is estimated according to the requirement that the probability of its associated texture class will be maximised. The estimation algorithm used in the system is simulated annealing. The set of $\beta_{i,i'}$ corresponds to the correlation value and thus represents the configuration of the pixels such that it can be classified as that texture class. In the system, this set of estimated β will be used as texture feature vector. It will be used as input of support vector machine such that the association between texture feature and texture class can be formed.

3.2 Texture learning using Support Vector Machine

Support vector machine have been widely used in recognition recently due to its non-linear classification power and thus be used to solve complicated recognition problem such as face recognition (e.g. [14]). Given data set: $\{(b_1, y_1), (b_2, y_2), ..., (b_l, y_l)\} \in B \times \{+1, -1\}$, support vector machine can learn to find out the association between b_i and y_i. In the proposed system, the b_i will be the texture feature set $\{\beta_{i,i'}\}$ after texture extraction on the input image and $\{+1, -1\}$ refers to edge and non-edge classes. During learning phase, the support vector machine will be trained. The classifier's parameters, α_i are learnt from data set $\{b_i, y_i\}$ under the criteria function, $max_\alpha \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j k(b_i, b_j)$. Gradient ascent approach is used in the system. During testing phase, the texture feature extracted from the image will be classified by the support vector machine. The determinant function can be written as $f(b) = sgn(\sum_{i=1}^{l} \alpha_i y_i k(b, b_i) + constant)$, where $k(\cdot, \cdot)$ is gaussian RBF kernel. The output will be an binary image with '1' indicates the edge class and '0' indicates the non-edge class. Mathematical details of support vector machine can be found in [10].

3.3 Texture Segmentation by Snake Fitting

Active contour [11] had been used in pattern location and tracking [15] for a long time. It is good at attaching to object with strong edge and irregular shape. The snake can be interpreted as parametric curve $v(s) = [x(s), y(s)]$.

In the proposed system, the initial position of the active contour is defined by the user. The active contour will move according to the refined energy function, $E_{snake}^* = \int_0^1 \{[E_{int}(v(s))] + [E_{texture}(v(s))] + [E_{con}(v(s))]\}ds$, where E_{int} represents the internal energy of the snake due to bending, $E_{texture}$ represents the texture-based image forces, and E_{con} represents the external constraint forces. The snake is said to be fitted if the E_{snake}^* is minimised.

The above equation is similar to commonly used snake equation but with the energy term $E_{texture}(v(s))$ replaces the original $E_{image}(v(s))$ which means image force. The energy term $E_{texture}(v(s))$ represents the energy of texture and is proportional to the negative of similarity of desired texture. This means the energy will be lower near the patch that shows desired texture (e.g. edge texture). Thus, the snake will search for strong edge in the binary texture map, that described in Section 3.2, along the direction toward the centroid of potential

region. It stops at the pixel with strong edge characteristic in the texture map. Thus, the term $E_{texture}(v(s))$ can be interpreted as the texture attractive force and the snake is texture sensitive. Texture represents a patch of pixels instead of a single pixel and texture-based analysis is more tolerant to noise compare with pixel-based analysis. Thus, texture is a much reliable feature than strong edge under pixel-based analysis.

3.4 Prediction by Kalman filtering

Kalman filtering [12] is a prediction-correction procedure. By encapsulating the motion of the object into internal states, Kalman filtering aims at finding appropriate states that gives best-fit observations. Dynamic equation and measurement equation will be used in Kalman filter for representing the change in internal states and conversion from internal state to observation respectively.

Since the motion of the vertebrae is planar, kalman filter is good enough to make prediction on the spinal motion. In the system, dynamic model is assumed to be common rigid body motion and the uni-modal Gaussian noise distribution is assumed. The state transition equation and the measurement equation used

$$
\text{are:} \quad
\begin{bmatrix} x_{t+1} \\ y_{t+1} \\ u_{t+1} \\ v_{t+1} \\ a_{t+1} \end{bmatrix}
=
\begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} x_t \\ y_t \\ u_t \\ v_t \\ a_t \end{bmatrix}
+ m_t, \text{ and }
\begin{bmatrix} x_t \\ y_t \end{bmatrix}
=
\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}
\begin{bmatrix} x_t \\ y_t \\ u_t \\ v_t \\ a_t \end{bmatrix}
+ n_t
$$

respectively, where x_t, y_t is the position of the landmark, u_t, v_t represent the velocity along x and y direction respectively, a_t represents the acceleration, and m_t, n_t represent the dynamic noise and measurement noise respectively. The equation for updating the parameters in Kalman filter can be found in [12].

4 Experiment and Result

The proposed system was implemented using Visual C++ under Microsoft Windows. The experiments were done on a P4 2.26 GHz computer with 512M Ram running Microsoft Windows.

4.1 Experiment 1: Low Contrast and Noisy Video fluoroscopic Image

In this experiment, the performance of the proposed feature learning and detection algorithm was evaluated. The vertebrae had to be segmented from the medical image with poor quality and low contrast. Actually, the image may not be segmented easily manually. There are totally 2 phases in this experiment. The first phase is learning phase where the texture pattern associated with edges is learnt and the second phase is vertebral boundary detection where the snake is fitted toward the edge detected.

In training phase, the total number of samples to be trained was around 1000, half of them was edge and half of them was non-edge. The samples were selected manually and were selected from images with similar illumination and contrast. The learning images and the testing images were randomly selected from the same video sequence and thus with similar illumination and contrast. In testing phase, the snake is initially mounted at location close to the vertebral boundary. The snake will then attach to the boundary automatically by using texture as heuristic.

The result of segmentation is showed in figure 2. It shows that the snake can fit the target vertebrae very well. The accuracy cannot be determined here due to no ground truth image provided. If the output is compare with the landmarks marked by a skilled physician, the relative root-mean-square error (the difference between the tracked corners and the physician-marked corners) is less than 3% in average out of 100 testing samples. The processing time is around 18s when the whole texture binary map is formed and the image with size 600 x 450 pixels. Some of the edge detection result of commonly used edge detector is shown in figure 3 for reference. It shows that the proposed method works much better than the commonly used edge detectors.

4.2   Experiment 2: Tracking spine in videofluoroscopic video

In this experiment, the performance of the whole system is evaluated. The system ran in the same way as described in section 2. Firstly, the accuracy on the reported location of the vertebra was tested. One of the vertebra in the video was tracked. The corners will be extracted and reported. Throughout a video sequence of 2 minutes, 200 sample frames was tested. The testing result is shown in figure 4(a). It shows that the reported location by the system and the marked location by the physician are very close. The relative root-mean-square error (the difference between the tracked location and the physician-marked location) is less than 5% in average. The processing time of each frame is aound 0.5 s because the edge pattern is now analysed along the snake instead of analysed the whole image.

The accuracy on intervertebral relation reported by the system were also tested. The angle between two vertebrae is usually used in most spinal motion analysis. Thus, the accuracy on the angle reported by the system were evaluated. The measurement methodology is shown in figure 5(a). The testing result is shown in figure 5(b). It shows the relative root-mean-square error (the angular difference between the tracked result and the physician-reported result) is quite large during initial phase but getting smaller after 30 frames. The relative root-mean-square error is lower than 10% in average in later stage.

Finally, the number of vertebrae that can be tracked by the system is evaluated. The result is shown in figure 4(b). It shows that totally four of the vertebrae, namely L2, L3, L4 and L5 can be tracked, provided that the illumination and the contrast is not varied a lot. The relative root-mean-square error reported is less than 10% in these four vertebrae.
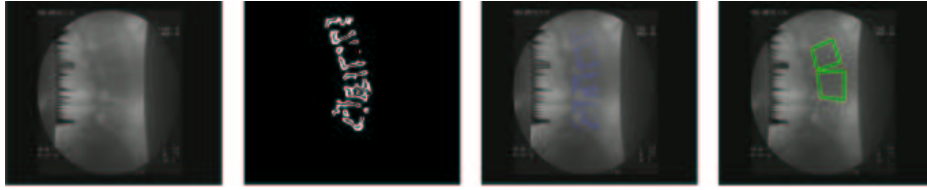
Fig. 2. *The first image shows the testing image. The second image shows the binary image after final classification. The third image shows the fused image constructed from the testing image and the binary image. The fourth image shows the snake attached to the boundary of the vertebrae.*

Fig. 3. *Edge detection result of some commonly used edge detector is shown.*

5 Conclusions

In this paper, a system for automatic spinal motion analysis is proposed. The proposed system requires less human intervention than common approaches by automating the edge detection and snake fitting. Operators may need to setup initial snake position in the first frame only. The edge will then be detected automatically using pattern recognition and the snake will fit toward the edge accordingly. The initial snake position in the next frame will be predicted through the use of dynamic that learnt from previous observations. Experimental results show that the proposed system can segment vertebrae from videofluoroscopic images automatically and accurately.

References

1. institute for Work & Health: Iwh fact sheet low-back pain. Internet resource, (Avaliable: http://www.iwh.on.ca/media/lowbackpain.php)
2. clinical Standards Advisory Group: Epidemiology review: the epidemiology and cost of back pain. HMSO. London (1994)
3. Breen, A.C., Allen, R., Morris, A.: A digital videofluoroscopic technique for spine kinematics. Journal of Medical Engineering and Technology **13** (1989) 109–113
4. Gatton, M.L., Pearcy, M.J.: Kinematics and movement sequencing during flexion of the lumbar spine. Clinical Biomechanics **14** (1999) 376–383
5. Cardan, C., Allen, R.: Measurement of spine motion for diagnosis of mechanical problems. Journal of Simulation Modeling in Medicine **1** (2000) 15–19
6. Page, W.H., Monteith, W., Whitehead, L.: Dynamic spinal analysis - fact and fiction. Chiropractic journal of Australia **23** (1993) 82–85
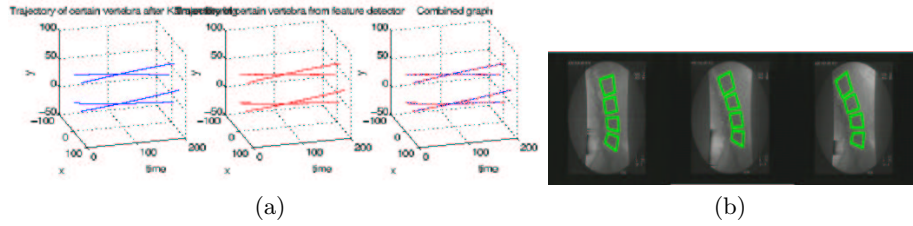
Fig. 4. *(a) The first two graphs show the reported location of the four vertebral corners along the time domain and time series data of location of the corresponding vertebral corners marked by physician respectively. The third graph combines the above two time series into one graph. (b) The tracking result of L2 to L5 vertebrae.*
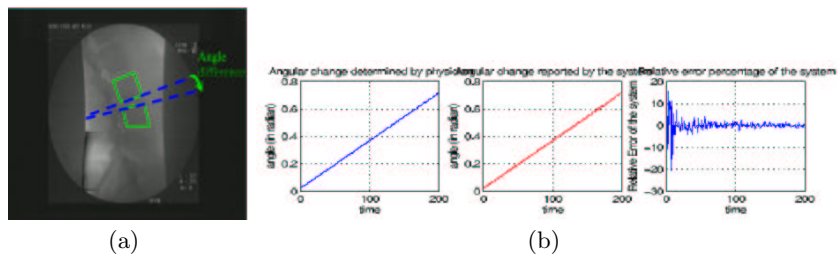

Fig. 5. *(a) The angle difference between middle two vertebrae is recorded in the experiment. (b) The left two graphs show the time series data on angle difference reported by the system and those measured by physician respectively. The third graph shows the corresponding relative root-mean-square error in percentage.*

7. Smyth, P.P., Taylor, C.J., Adams, J.E.: Automatic measurement of vertebral shape using active shape models. Image and Vision Computing **15** (1997) 575–581
8. Zheng, Y., Nixon, M.S., Allen, R.: Automated segmentation of lumbar vertebrae in digital videofluoroscopic images. IEEE Trans. Medicial Imaging **23** (2004) 45–52
9. Chellappa, R., Jain, A.: Markov Random Fields: Theory and Applications. Academic Press (1993)
10. Scholkopf, B., Burges, C.J.C., Smola, A.J.: Advances in Kernel Methods - Support Vector Learning. MIT Press, Cambridge, MA (1999)
11. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. In: Proc. Int. Conf. on Computer Vision. (1987) 259–268
12. Kalman, R.E.: A new approach to linear filtering and prediction problems. Trans. ASME - Journal of Basic Engineering **82(Series D)** (1960) 35–35
13. Krishnamachari, S., Chellappa, R.: Multiresolution gauss-markov random-field models for texture segmentation. IP **6** (1997) 251–267
14. Fransens, R., DePrins, J., Gool, L.J.V.: Svm-based nonparametric discriminant analysis, an application to face detection. In: ICCV03. (2003) 1289–1296
15. Blake, A., Isard, M.: Active Contours. Springer (1998)