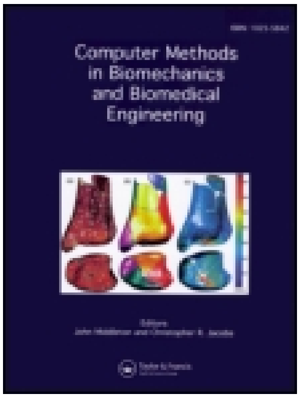
This article was downloaded by: [University of Waterloo]
On: 11 July 2011, At: 12:43
Publisher: Taylor & Francis
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Computer Methods in Biomechanics and Biomedical Engineering

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gcmb20

A systematic approach to feature tracking of lumbar spine vertebrae from fluoroscopic images using complex-valued wavelets

Alexander Wong [a], Nadine M. Dunk [b] & Jack P. Callaghan [b]

[a] Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada

[b] Department of Kinesiology, University of Waterloo, Waterloo, Ontario, Canada

Available online: 30 Sep 2009

To cite this article: Alexander Wong, Nadine M. Dunk & Jack P. Callaghan (2009): A systematic approach to feature tracking of lumbar spine vertebrae from fluoroscopic images using complex-valued wavelets, Computer Methods in Biomechanics and Biomedical Engineering, 12:5, 607-616

To link to this article: http://dx.doi.org/10.1080/10255840902802891

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan, sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Computer Methods in Biomechanics and Biomedical Engineering
Vol. 12, No. 5, October 2009, 607–616

Taylor & Francis
Taylor & Francis Group

A systematic approach to feature tracking of lumbar spine vertebrae from fluoroscopic images using complex-valued wavelets

Alexander Wong[a]*, Nadine M. Dunk[b] and Jack P. Callaghan[b]

[a]*Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada;* [b]*Department of Kinesiology, University of Waterloo, Waterloo, Ontario, Canada*

(*Received 10 September 2008; final version received 6 February 2009*)

This paper presents a systematic approach to lumbar spine vertebrae tracking in fluoroscopic images using complex-valued wavelets. The proposed algorithm is designed specifically based on a set of performance criteria associated with the detection and tracking of feature points in lumbar spine vertebrae from fluoroscopic images. The algorithm handles contrast and illumination non-homogeneities and noise in fluoroscopic images through the use of local phase information obtained using complex-valued wavelets. The algorithm is capable of tracking feature points that undergo various geometric deformations caused during the fluoroscopic imaging process by defining a descriptor that is invariant to scale and rotation and robust to affine, projective and mild pin-cushion distortions. The algorithm has been tested using dynamic sagittal fluoroscopic videos of the lumbar-sacral region and testing results indicate that the algorithm achieves good tracking performance of lumbar spine vertebrae in fluoroscopic images that exhibit contrast and illumination non-homogeneities as well as noise, with mean root mean square error of less than 0.40 mm under in all test sequences.

Keywords: tracking; lumbar spine; vertebrae; fluoroscopic images; complex-valued wavelets

Downloaded by [University of Waterloo] at 12:43 11 July 2011

1. Introduction

One of the most common and widespread problems associated with industrialisation is low back pain. The factors that contribute to low back pain are not well understood, making it difficult to provide good clinical diagnosis on this condition. Abnormal lumbar intersegmental motion has been identified as a potential culprit for the development and persistence of low back pain (Breen et al. 1989; Okawa et al. 1998; Teyhen et al. 2007). Therefore, it is important to study the dynamic intersegmental motion patterns of the lumbar spine to better understand the symptoms associated with low back pain to improve diagnosis and treatment.

In recent years, one of the most effective methods of capturing the motion of the lumbar spine is through the use of fluoroscopic imaging. In traditional radiographic systems, the number of images that can be acquired over time is highly limited due to the level of radiation exposure to the patient. As such, images captured using these traditional systems are often limited to the neutral and extreme positions of the lumbar spine motion. This makes it very difficult to analyse mid-range motion characteristics of the lumbar spine, where abnormal movement may occur (Teyhen et al. 2007). Fluoroscopic imaging systems incorporate an X-ray source, an image intensifier and a video camera, and are capable of acquiring a series of images. The distinguishing component of a standard fluoroscopic system is the image intensifier. Image intensifiers are far more sensitive than a standard 400-speed screen-film cassette and can produce images using less radiation (Cholewicki et al. 1991; Bushberg et al. 2002). Thus, fluoroscopy is an appealing alternative to traditional radiography for studying lumbar spine motion. There are many issues associated with lumbar spine vertebrae tracking in fluoroscopic images, such as distortion, low contrast resolution, and signal non-homogeneity, that need to be addressed as a whole, which will be discussed later in the paper. As such, a tracking system that is robust to distortion, low contrast resolution, signal non-homogeneity, and noise is greatly desired. The goal of this proposed algorithm is to address all of these issues, which current methods have not been able to successfully address in their entirety.

The main contribution of this paper is a systematic approach to lumbar spine vertebrae tracking in fluoroscopic images using complex-valued wavelets. The intended use of the proposed lumbar spine vertebrae tracking system is for post-imaging analysis to study the motion patterns of the lumbar spine. The algorithm is designed specifically to satisfy a comprehensive list of performance criteria that are important to tracking vertebrae features. The proposed algorithm employs techniques based on complex-valued wavelets in the detection and representation processes to address contrast and illumination non-homogeneities and noise that may affect tracking accuracy in the series of fluoroscopic images.

*Corresponding author. Email: a28wong@engmail.uwaterloo.ca

ISSN 1025-5842 print/ISSN 1476-8259 online
© 2009 Taylor & Francis
DOI: 10.1080/10255840902802891
http://www.informaworld.com

The identified vertebrae feature points and their associated descriptors are designed to be distinctive and have a high probability of being tracked over time. Robustness to various geometric deformations in fluoroscopic images is achieved by defining a local descriptor that is scale and rotation invariant, as well as robust to affine, projective and mild pin-cushion transformations. A fast and robust descriptor matching scheme is employed in the proposed method to reduce the performance overhead associated with matching vertebrae feature points. Finally, a feature point rejection schemed based on maximum distance sample consensus (MDSAC) is used to provide an efficient and effective method of reducing the effect of erroneous vertebrae feature point matches on tracking accuracy.

In this paper, issues associated with tracking lumbar spine vertebrae in fluoroscopic images are discussed in Section 2. Previous work related to lumbar spine vertebrae tracking is discussed in Section 3. The set of performance criteria is described in Section 4. The theory underlying the design process of each step in the proposed vertebrae tracking algorithm is presented and explained in Section 5. The instrumental and experimental setup is presented in Section 6. The proposed algorithm is validated using dynamic sagittal fluoroscopic videos of the lumbo-sacral region and the results are presented in Section 7. Finally, conclusions are drawn based on the results in Section 8.

2. Issues with lumbar spine vertebrae tracking in fluoroscopic images

There are many issues that make feature tracking of lumbar spine vertebrae in fluoroscopic images a particularly challenging problem to solve. As mentioned before, fluoroscopy systems are equipped with an image intensifier in line with the X-ray source. Image intensifier technology requires electrons to be focused using an input screen with a curved surface, thus resulting in 'pin cushion' distortion in the output image (Cholewicki et al. 1991; Bushberg et al. 2002).

Fluoroscopic images often exhibit contrast and illumination non-homogeneities. Due to the lower exposure levels used in fluoroscopy, the contrast resolution is lower than traditional radiographic systems. Therefore, fluoroscopic imaging systems produce images with relatively low signal to noise ratios. While contrast resolution can be increased when higher exposure rates are used (by increasing the X-ray current and/or voltage), this has the major disadvantage of increased higher patient radiation dosage (Bushberg et al. 2002). Furthermore, the size or thickness of the patient will also affect the quality of the output image. This is because thicker regions attenuate more radiation, thus resulting in less radiation from striking the image intensifier and generating less illumination (Bushberg et al. 2002). Therefore, different regions of tissue thickness and density will result in the brightness of the same image content to vary spatially across an individual frame, or temporally from frame to frame.

Another issue associated with tracking lumbar spine vertebrae features in fluoroscopic images is that the motion of the patient can lead to minor viewpoint variations between adjacent frames in a lumbar spine vertebrae sequence. Furthermore, the motion of the lumbar spine will lead to the motion of other internal anatomical structures (such as the pelvis, or intestines). Consequently, this can result in the visual occlusion and overlay of lumbar spine vertebrae and anatomical features thereby make it difficult to track features over time.

3. Previous work

Various methods have been proposed for the purpose of tracking lumbar spine vertebrae in fluoroscopic images (Muggleton and Allen 1997; Bifulco et al. 2001; Wong et al. 2004; Zheng et al. 2004; Penning et al. 2005; Wong et al. 2006). In several studies (Muggleton and Allen 1997; Bifulco et al. 2001; Penning et al. 2005), a template matching approach to lumbar spine vertebrate tracking based on the cross-correlation cost function was utilised. There are two major drawbacks to the aforementioned approach. First, these methods use rigid templates for the lumbar spine vertebrae, making them sensitive to geometric distortions, which may occur to a certain degree in the outer edges of the fluoroscopic image. Second, the use of the cross-correlation cost function, while robust against uniformly linear contrast variations, is highly sensitive to illumination and contrast non-homogeneities. This can be problematic in the case of fluoroscopic imaging, where such non-homogeneities can occur both spatially and temporally. Wong et al. (2004, 2006) used a learning-based approach, where texture patterns modelled by Markov random fields (MRFs) are learned using support vector machines (SVMs). An active contour was then fitted around the edge using the texture information and then tracked over time using a Kalman filter (Kalman 1960). There are two main limitations with this approach. First, the use of a learning-based approach with active contours can be computationally expensive, particularly for analysing long fluoroscopic sequences. Furthermore, since it directly relies on intensity in modelling texture patterns, it is sensitive to contrast and illumination non-homogeneities.

Few methods have defined a comprehensive set of performance criteria for what constitutes an effective method for tracking lumbar spine vertebrae in fluoroscopic images. This lack of performance criteria formalisation

Downloaded by [University of Waterloo] at 12:43 11 July 2011

Downloaded by [University of Waterloo] at 12:43 11 July 2011

is evident in the individual drawbacks exhibited by existing techniques. The formalisation of performance criteria is important as it allows for a systematic approach to the design of an algorithm that addresses specific issues that pertain to a problem. A prime example of how the formalisation of performance criteria can aid in the design of an effective algorithm is the edge detection algorithm proposed by Canny (1986), which has been very successful and widely used since its introduction. Therefore, the proposed algorithm presented in this paper utilises a systematic approach to lumbar spine vertebrae tracking based on a comprehensive set of performance criteria.

4. Performance criteria

As mentioned earlier, there are a number of issues associated with tracking lumbar spine vertebrae in fluoroscopic images. To address these issues in a systematic manner, performance criteria can be established to tailor the design of the tracking system specifically for lumbar spine vertebrae tracking. Several important performance criteria can be defined as follows:

(1) *Accuracy*: The identified vertebrae feature points should be highly distinctive and have a high probability of being tracked from frame to frame. There should also be a low probability of falsely detecting feature points that appear distinctive due to noise, since such points can lead to erroneous feature point matches. This is particularly important in the case of fluoroscopic images, which are often characterised by low signal to noise ratios due to low exposure levels. Furthermore, the points identified as vertebrae feature points should be within 0.50 mm of the centre of the true anatomical feature, which we will consider to an acceptable error for an accurate study of lumbar spinal motions given that the mean error of manual digitising on spinal fluoroscopic images is up to 10% of the translational range of motion (Harada et al. 2000), or approximately 0.50 mm in the case of a seated forward flexion motion (Takayanagi et al. 2001). Finally, the vertebrae feature points should also be as consistent as possible between different frames.

(2) *Robustness*: Over time (or from frame to frame), the appearance of the lumbar spine vertebrae and the internal anatomical structures often change due to factors such as illumination and contrast non-homogeneity, environment noise due to soft issue scatter, changes in orientation and scale, occlusion, and changes in viewpoint. Therefore, it is important that the vertebrae feature points found in one frame are matched with the corresponding vertebrae feature points found in subsequent frames irrespective of such changes.

(3) *Efficiency*: Since the system is designed for post-imaging analysis, computational efficiency is not critical for proper system functionality. However, given that there are many images within each motion sequence acquired, it is still important that the computational overhead required to perform the tracking process is as low as possible while maintaining tracking accuracy in order to obtain lumbar spinal motion results in a timely manner. Furthermore, it is also important to minimise the amount of user intervention required to perform the tracking process.

The proposed algorithm attempts to satisfy all of the aforementioned performance criteria in hopes of achieving improved tracking performance of lumbar spine vertebrae in fluoroscopic images.

5. Proposed method

The proposed algorithm utilises several techniques to address the specific issues and challenges associated with tracking lumbar spine vertebrae features in fluoroscopic images. As such, it is intuitive to use a systematic approach to describe the design of the proposed algorithm. Therefore, for each step of the proposed algorithm, the key performance criteria that apply to the corresponding step are discussed, the drawbacks associated with existing techniques are described, and the techniques used to address these drawbacks are explained in detail.

5.1 *Vertebrae feature point detection*

The first step in the proposed algorithm is to detect a set of feature points pertaining to lumbar spine vertebrae from the current frame and a subsequent frame, respectively. Various algorithms have been proposed for the purpose of feature point detection (Harris and Plessey 1988; Kovesi 2003; Fauqueur et al. 2006). However, there are limitations to these techniques that make it difficult to address the aforementioned challenges to vertebrae feature point detection in fluoroscopic images. Commonly-used feature detection algorithms such as those proposed by Harris and Plessey (1988) and Fauqueur et al. (2006), and the DoG maxima method (Lowe 2004) are highly sensitive to illumination and contrast non-homogeneity that are often found in fluoroscopic images. Furthermore, many of these techniques are highly sensitive to noise and utilise Gaussian pre-filtering to suppress noise, which can also significantly reduce the distinctiveness of vertebrae interest points. This is especially problematic given the low contrast resolution of fluoroscopic images.

A more suitable approach is that proposed by Kovesi (2003), which attempts to address the issue of illumination and contrast non-homogeneity by using local frequency characteristics obtained from complex-valued wavelets. By utilising only the phase information, these detection methods are largely invariant to illumination and contrast non-homogeneity. Furthermore, this approach has been shown to be highly robust to noise without the need for pre-filtering, provide improved localisation and account for variations due to orientation. These benefits make it better suited for detecting vertebrae feature points in fluoroscopic images. However, the technique proposed by Kovesi (2003) does not provide the information that distinguishes feature points detected at different scales. This is an important factor that needs to be accounted for due to the scale distortions caused by patient motion. As such, the proposed algorithm extends upon the robust feature point detection method described by Kovesi (2003) such that the dominant scale information regarding each vertebrae feature point can be obtained from each fluoroscopic image.

The proposed vertebrae feature point detection scheme can be described as follows. Local frequency information is extracted at each pixel over multiple scales and orientations using directional over-complete complex-valued wavelets such as Gabor wavelets and dual-tree complex wavelets (Selesnick et al. 2005). The local amplitude and phase for a given point $\underline{x}$ at wavelet scale n can be computed as follows:

$$A_n(\underline{x}) = \sqrt{\left(I(\underline{x}) \times F_n^e\right)^2 + \left(I(\underline{x}) \times F_n^o\right)^2}, \tag{1}$$

$$\phi_n(\underline{x}) = \tan^{-1}\left(\frac{\left(I(\underline{x}) \times F_n^e\right)}{\left(I(\underline{x}) \times F_n^o\right)}\right), \tag{2}$$

where F_n^e and F_n^o are the even- and odd-symmetric wavelets at scale n. Using this information, the local phase coherence at a point in the image $\underline{x}$, orientation θ, and over a range of N scales can be defined as follows:

$$\rho(\underline{x}, \theta) = \frac{\sum_n^N W(\underline{x}, \theta) \lfloor A_n(\underline{x}, \theta) \Delta\Phi(\underline{x}, \theta) - T \rfloor}{\sum_n A_n(\underline{x}, \theta) + \varepsilon}, \tag{3}$$

$$\begin{aligned} \Delta\Phi(\underline{x}, \theta) = {} & \cos\left(\phi_n(\underline{x}, \theta) - \bar{\phi}(\underline{x}, \theta)\right) \\ & - \left|\sin\left(\phi_n(\underline{x}, \theta) - \bar{\phi}(\underline{x}, \theta)\right)\right|, \end{aligned} \tag{4}$$

where W represents the frequency spread weighting factor, A_n and ϕ_n represent the amplitude and phase at wavelet scale n, respectively, $\bar{\phi}$ represents the weighted mean phase, T represents the noise threshold and ε is a small constant used to avoid division by zero. The parameters used are the same as those described in Kovesi (2003). Based on this formulation, as the complex-valued wavelet coefficients come into phase, local phase coherence approaches one (assuming non-zero amplitudes).

As the local phase coherence can vary as a result of orientation, the vertebrae feature significance of a point can be defined as the minimum moment of local phase coherence $\mu(\underline{x})$ over multiple orientations:

$$\frac{1}{2}\sum_\theta \rho(\underline{x}, \theta)^2 + \frac{1}{2}\sqrt{\begin{aligned} & 4\left(\sum_\theta (\rho(\underline{x}, \theta)\sin(\theta))(\rho(\underline{x}, \theta)\cos(\theta))\right)^2 \\ & -\left(\sum_\theta \left[(\rho(\underline{x}, \theta)\cos(\theta))^2 - (\rho(\underline{x}, \theta)\sin(\theta))^2\right]\right)^2 \end{aligned}} \tag{5}$$

where $\rho(x, \theta)$ is the local phase coherence at orientation θ. The minimum moment of local phase coherence increases as the vertebrae feature significance increases.

Once the minimum moment of local phase coherence has been determined, the vertebrae feature points are located at the local minimum moment maxima. To avoid the computational overhead associated with representing and matching for large numbers of feature points, it is necessary to restrict the number of vertebrae feature points selected from each frame to a reasonable level. As such, a method for selecting a fixed number of feature points is desired. This is accomplished in the proposed algorithm by retaining the k highest local minimum moment maxima. Based on testing, setting $k = 100$ was found to provide good results. Finally, to refine the location of the feature points to achieve sub-pixel accuracy, a 2D quadratic function is fitted to the minimum moment of local phase coherence within a neighbourhood around each feature point and the interpolated location of the maximum is determined as the final location of the feature point.

Given the set of vertebrae feature points, the dominant scale, s, at which the feature point is most distinctive is determined as the minimum moment maxima of local phase coherence at the location of the feature point over a range of scales:

$$s(\underline{x}) = \arg\max_s(\mu_s(\underline{x})), \tag{6}$$

where $\mu_s(\underline{x})$ is the minimum moment of local phase coherence at location $\underline{x}$ and scale s. Some sample feature points are shown in Figure 1.

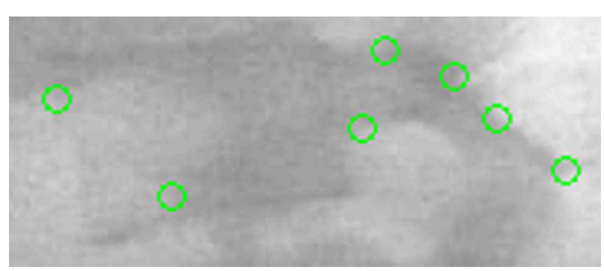


Figure 1. Sample feature points detected on a vertebra.

Downloaded by [University of Waterloo] at 12:43 11 July 2011

Downloaded by [University of Waterloo] at 12:43 11 July 2011

5.2 *Vertebrae feature point description*

The second step in the proposed vertebrae tracking algorithm is to create a descriptor for each vertebrae feature point identified to represent the vertebrae feature characteristics from the current frame and a subsequent frame, respectively. Various descriptors have been proposed for the purpose of feature point description (Schiele and Crowley 1996; Schmid and Mohr 1997; Lowe 2004; Wong and Clausi 2007). Most recent feature point descriptors are invariant to similarity transformations such as scale and rotation, as well as robust to various affine distortions. Nevertheless, there are limitations to these techniques that make it difficult to address the aforementioned challenges to feature point description as applied to vertebrae tracking in fluoroscopic images. The descriptor proposed by Schmid and Mohr (1997) is highly sensitive to illumination and contrast non-homogeneities, making it ill-suited for representing information in fluoroscopic images. Both Schiele and Crowley (1996) and Lowe (2004) attempted to address this issue by using normalisation techniques. The main problem with this approach is that normalisation has a tendency to amplify noise, making the descriptors highly sensitive to the presence of noise. This is problematic given the low signal to noise ratios that characterise fluoroscopic images. Therefore, these descriptors require pre-filtering to suppress noise, which can also significantly reduce the distinctiveness of the vertebrae characteristics. A more recent technique was proposed by Wong and Clausi (2007), which addressed the issue of illumination and contrast non-homogeneity by using local frequency characteristics obtained from complex-valued wavelets. By utilising only the phase information, the resulting descriptors are largely invariant to illumination and contrast non-homogeneity. Unfortunately, this descriptor is not scale and orientation invariant, and is sensitive to localisation errors in the actual feature point. Therefore, a descriptor that is invariant to illumination and contrast non-homogeneity as well as robust to noise and various geometric distortions is desired for the purpose of vertebrae representation.

The proposed vertebrae feature point description scheme can be described as follows. Based on the local phase coherence information extracted using complex-valued wavelets during the feature point detection step, the structural significance pertaining to a point can be defined as the maximum moment of local phase coherence $M(\underline{x})$ over multiple orientations:

$$\frac{1}{2}\sum_{\theta}\rho(\underline{x},\theta)^2+\frac{1}{2}\sqrt{\begin{array}{c}4\left(\sum_{\theta}(\rho(\underline{x},\theta)\sin(\theta))(\rho(\underline{x},\theta)\cos(\theta))\right)^2\\ +\left(\sum_{\theta}\left[(\rho(\underline{x},\theta)\cos(\theta))^2-(\rho(\underline{x},\theta)\sin(\theta))^2\right]\right)^2\end{array}} \tag{7}$$

where $\rho(\underline{x},\theta)$ is the local phase coherence at orientation θ. According to this formulation, the maximum moment of local phase coherence increases as the structural significance increases. Therefore, maximum moment of local phase coherence can be used to represent vertebrae structural characteristics.

Once the maximum moment of local phase coherence has been determined, the maximum moment gradient orientation can be defined as follows:

$$\theta(x,y)=\tan^{-1}\left(\frac{(\partial M(x,y)/\partial y)}{(\partial M(x,y)/\partial x)}\right). \tag{8}$$

To determine the dominant orientation pertaining to a particular vertebrae feature point, an orientation histogram is constructed based on the maximum moment distribution within a neighbourhood around the feature point at its dominant scale. Each entry into the orientation histogram is weighted by the product of the corresponding maximum moment and a 2D Gaussian function centred at the location of the feature point. Similar to the approach taken in Lowe (2004), any local peaks in the orientation histogram that are within 80% of the maximum peak is selected as a dominant orientation and a vertebrae feature point descriptor is created for each of these dominant orientations. This is done to reduce the effect of erroneous dominant orientation information on the feature point matching process.

The local descriptor pertaining to a particular vertebrae feature point can then be constructed based on the maximum moment distribution of neighbouring points with respect to maximum moment gradient orientation. Similar to the dominant orientation selection process, a Gaussian weighted orientation histogram is constructed based on the maximum moment distribution within a neighbourhood around the feature point at its dominant scale. The Gaussian weighting reduces the effect of localisation errors in the actual feature point on the local descriptor. Furthermore, the orientation histogram is normalised by the number of sample points within a neighbourhood such that descriptors can then be compared between feature points of different scales. For each dominant orientation pertaining to a vertebrae feature point, a local descriptor is created by performing a circular shift on the aforementioned orientation histogram relative to the dominant orientation. This vertebrae feature point description scheme is illustrated in Figure 2. The proposed feature point descriptor is relatively simple to construct and also allows for fast feature point matching, which is important in situations where a large number of feature points are needed.

5.3 *Vertebrae feature point matching*

The third step in the proposed algorithm is to match vertebrae feature points from the current frame to the

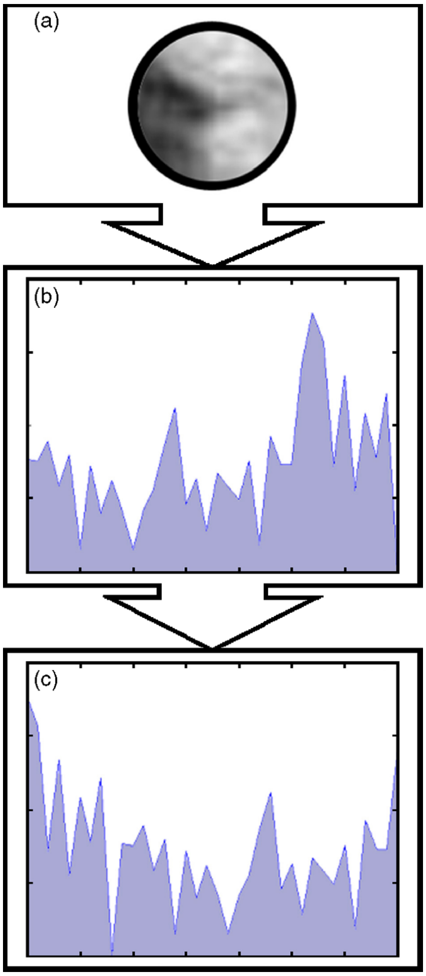


Figure 2. Vertebrae feature point description process: for a given neighbourhood around a vertebrae feature point at a given scale (a), a weighted orientation histogram is created (b). A circular shift is performed based on the dominant orientation (c).

vertebrae feature points from a subsequent frame in the fluoroscopic sequence. Various similarity metrics have been used for the purpose of feature point matching. One of the most commonly used similarity metrics for vertebrate feature matching is the cross-correlation coefficient (Muggleton and Allen 1997; Bifulco et al. 2001; Penning et al. 2005). The use of the cross-correlation coefficient to compare intensity values, as done by existing methods, is very sensitive to contrast and illumination non-homogeneity. Even if used to compare the proposed invariant vertebrae descriptor, the cross-correlation coefficient still runs into problems as it is relatively sensitive to outliers. This is problematic given the fact that fluoroscopic images are characterised by low contrast resolution, low signal to noise ratio, and minor geometric distortions which can result in outliers. To reduce the influence of outliers on the vertebrae feature point matching process, a Pearson Type VII distance metric is utilised:

$$d(f,g) = \sum_{\theta} \ln\left(\sqrt{1 + \left(O_f(\theta) - O_g(\theta)\right)^2}\right), \qquad (9)$$

where f and g are vertebrae feature points from the current frame and a subsequent frame, respectively, O_f and O_g the local descriptors that correspond with feature points f and g, respectively, and θ the maximum moment gradient orientation. The Pearson Type VII distance metric is a redescending estimator and so the influence of outliers on this metric tends to zero. Furthermore, it is relatively efficient to compute, particularly for a compact feature vector such as the proposed vertebrae descriptor. To further improve computational overhead, the search space is restricted to a radius of 20% of the smallest frame dimension. This is a reasonable restriction for the purpose of lumbar spine vertebrae tracking as the motion can be assumed to be relatively small from frame to frame given the physical limitations of the lumbar spine.

5.4 *Vertebrae feature point rejection*

The final step in the proposed algorithm is to prune erroneous vertebrae feature point matches from the set of matched feature point pairs. Various algorithms have been used for the purpose of feature point rejection purposes. Some of the most widely used feature point rejection techniques are those that utilise data resampling. These include the random sample consensus (RANSAC) algorithm (Fischler and Bolles 1981), least median of squares (LMS) (Stewart 1999) and more recently the MDSAC (Wong and Clausi 2007). Such techniques are popular for several reasons. First, no *a priori* knowledge regarding the distribution of feature points is required, which makes them well suited for a wide range of applications. In addition, these techniques are relatively simple and can be implemented in an efficient manner. Finally, these techniques have proven to be very effective for identifying outlier measurements.

The proposed rejection scheme is based on the MDSAC algorithm, which is a variant of RANSAC that has been shown to provide good robustness in situations characterised by high outlier-to-inlier ratios while achieving faster convergence than standard RANSAC. The algorithm used to prune erroneous vertebrae feature point pairs in the proposed algorithm can be described as follows. First, feature point pairs are drawn randomly from the pool of feature point pairs to form multiple data sets. The cumulative Euclidean distances between the feature point pairs within each data set are computed. The data set with the maximum cumulative Euclidean distance is selected and used to estimate the motion model of the

Downloaded by [University of Waterloo] at 12:43 11 July 2011

target objects. The number of inlier feature point pairs is then determined based on a root mean square error (RMSE) distance between the actual feature point location and the estimated feature point location. This process is repeated over K iterations and the estimated motion model that yields the greatest number of inlier feature point pairs is selected. The number of iterations is determined as follows:

$$K = \frac{\log(1-p)}{\log\left(1-(n_{\mathrm{inliers}}/N)^s\right)}, \qquad (10)$$

where p is the desired probability of selecting at least one motion model that is free of outliers within K iterations, n_{inliers} the number of feature point pairs that satisfy the current estimated motion model, N the total number of feature point pairs and s the minimum number of feature point pairs necessary to estimate the motion model. The feature point pairs that do not satisfy the estimated vertebrae motion model are pruned from the set of feature point pairs. The remaining inlier vertebrae feature point pairs can then be used to refine the estimated vertebrae motion model.

6. Instrumental and experimental setup

To investigate the motion of the lumbar spine, videos were obtained using a fluoroscope system while the participants moved their spine from an upright-seated posture to a slouched seated posture. The video fluoroscope (Siremobil Compact (L) Mobile X-ray Image Intensifier system, Siemens Medical Solutions USA, Inc., Malvern, PA, USA) was equipped with a 9-inch. image intensifier and a source to image intensifier distance of 100 cm. The average X-ray technique factors were 3.4 mA and 107 kV and the total imaging time did not exceed 20 s. The total effective dose was less than 6 mSv, which was less than that of a typical CT examination. The FDA also estimates that 10 mSv of X-ray exposure increases the risk of fatal cancer by 0.05%, which is considered insignificant. The video feed from the fluoroscope was captured at 30 Hz directly to a computer hard drive using a digital video capture device and software (DVD Xpress DX2, ADS Technologies, Cerritos, CA, USA). To evaluate the effectiveness of the proposed algorithm, an individual vertebra was tracked over time in six different test dynamic sagittal fluoroscopic videos (acquired from three males and three females) of the lumbar-sacral region ranging from the top of the sacrum to the top of the third lumbar vertebra. The resolution of each of the test sets is 0.222 mm. A summary of the test sets is shown below:

- Male1: male (age = 22; height = 1.80 m; mass = 77.6 kg)
- Male2: male (age = 28; height = 1.71 m; mass = 69.0 kg)
- Male3: male (age = 28; height = 1.76 m; mass = 70.0 kg)
- Female1: female (age = 22; height = 1.68 m; mass = 49.9 kg)
- Female2: female (age = 25; height = 1.70 m; mass = 59.0 kg)
- Female3: female (age = 24; height = 1.57 m; mass = 45.8 kg)

A sample frame from each sequence is shown in Figure 3. To evaluate the tracking performance of the proposed algorithm, the RMSE (in mm) was computed between the estimated position of detected feature points for the vertebra and the ground-truth locations of these feature points manually selected by an expert over the video sequences. For comparison purposes, the tracking methods proposed by Penning et al. (2005) and Bifulco et al. (2001) were also evaluated using the test sets. All tested methods were implemented in MATLAB, with all tests were performed on an Intel Core 2 Duo 1.67 GHz PC with 2 GB of RAM. Due to the randomness of the vertebrae feature point rejection process, a total of 30 trials was performed for each test set and the minimum, mean and maximum RMSE for each test set is reported. The test sets used to validate the proposed method exhibited low contrast resolution and were characterised by contrast and illumination non-uniformity. Furthermore, the data sets were also contaminated by high levels of noise due to soft tissue scatter and the differing densities of the tissues in the abdomen. Finally, due to the motion of the lumbar spine as well as the surrounding anatomical objects, partial occlusion also occurred. All these characteristics inherent to the test data sets make them well-suited for testing robustness against different issues associated with target representation and localisation. For test purposes, the proposed method was configured such that the maximum number of feature points found during the detection step was 100 points for each frame.

7. Results

The RMSE between the estimated locations of corresponding vertebrae feature points over the frames of the video sequences and the ground-truth locations for each of the tested methods is summarised in Table 1. It can be observed that the mean RMSE for the proposed algorithm was noticeably lower than the other tested methods for all test fluoroscopic sequences. The mean RMSE of the proposed method was below 0.40 mm for all test fluoroscopic sequences, which satisfies the accuracy criteria of 0.50 mm. The peaks in RMSE, as indicated by the maximum RMSE, generally occur in situations where there is fast lumbar spine motion, resulting in large jumps between the position of the vertebrae. Nevertheless, the maximum RMSE for all test fluoroscopic sequences are

Downloaded by [University of Waterloo] at 12:43 11 July 2011

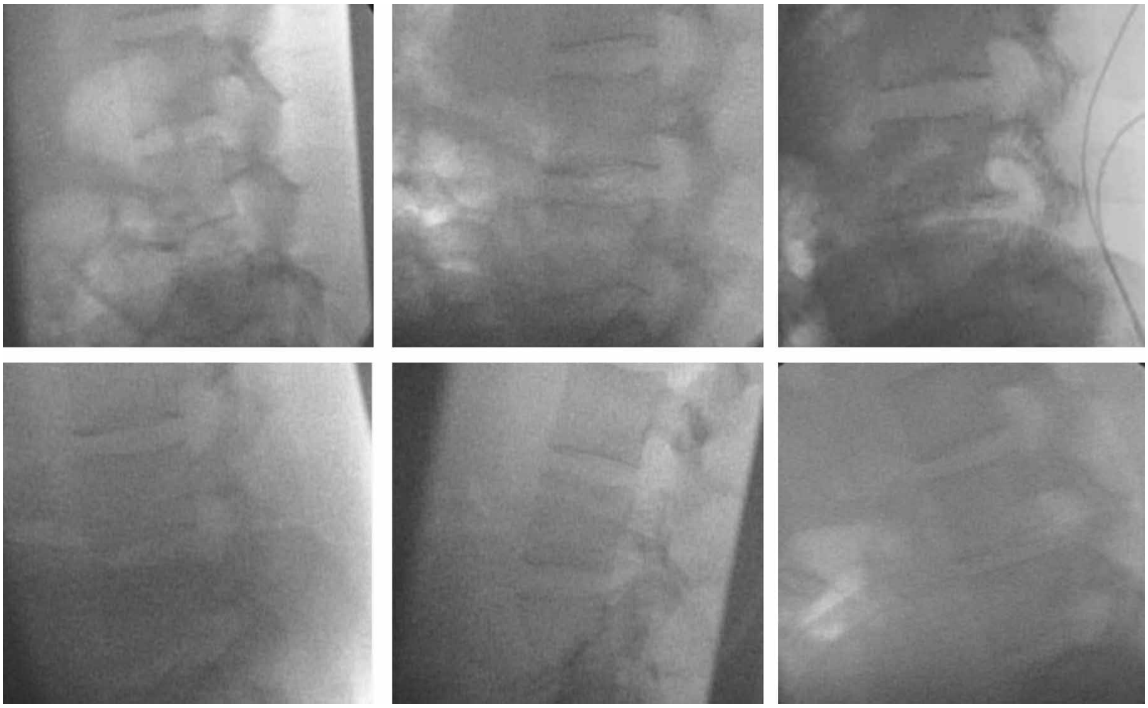


Figure 3. A sample frame from each test sequence (clockwise from top-left): (a) Male1, (b) Male2, (c) Male3, (d) Female1, (e) Female2 and (f) Female3.

reasonable low, demonstrating the robustness of the proposed method in situations of fast lumbar spine motion.

Sample tracking results for the Female3 and Male1 sequences are shown in Figures 4 and 5, respectively. The white square indicates the estimated location and orientation of the tracked vertebrae. By visual inspection, it can be seen that the estimated location and orientation of the tracked lumbar spine vertebrae is accurately tracked from one frame to the next. This demonstrates that the proposed algorithm is capable of satisfying the robustness criteria for various conditions inherent in fluoroscopic images.

Finally, the average processing time of an individual video frame for all test fluoroscopic sequences is summarised in Table 1. It can be observed that the average processing time of the proposed method is 47.25 s/frame, which is noticeably lower than the methods proposed by Penning et al. (2005) and by Bifulco et al. (2001). This demonstrates that the proposed algorithm is capable of satisfying the efficiency criteria. As such, the experimental results show that the proposed method satisfies all three of the defined performance criteria (Table 2).

8. Conclusions and future work

In this paper, we have introduced a systematic approach to the problem of lumbar spine vertebrae tracking in fluoroscopic images. Specific issues and challenges associated with detecting and tracking vertebrae features

Table 1. Tracking accuracy for fluoroscopic images.

	RMSE (mm)								
	Penning et al. (2005)			Bifulco et al. (2001)			Proposed		
Test set	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
Male1	0.19	0.95	1.56	0.28	0.84	1.23	0.11	0.29	0.49
Male2	0.35	0.76	1.02	0.34	0.73	0.97	0.15	0.24	0.35
Male3	0.39	0.73	0.95	0.23	0.66	0.85	0.12	0.28	0.36
Female1	0.22	1.14	1.96	0.35	0.92	1.67	0.17	0.38	0.65
Female2	0.34	0.70	0.94	0.32	0.85	1.01	0.15	0.34	0.46
Female3	0.47	0.96	1.25	0.19	0.87	0.92	0.15	0.32	0.56

Downloaded by [University of Waterloo] at 12:43 11 July 2011

Downloaded by [University of Waterloo] at 12:43 11 July 2011

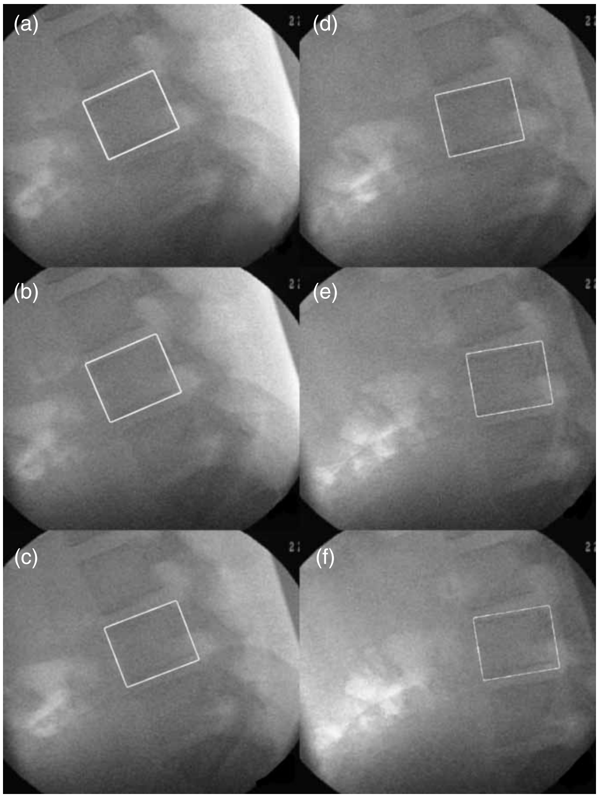


Figure 4. Example tracking results for the Female3 sequence. The frames are in alphabetical order. The white square indicates the estimated location and orientation of the tracked vertebrae. It can be observed that the estimated location and orientation is accurate across the entire video sequence.

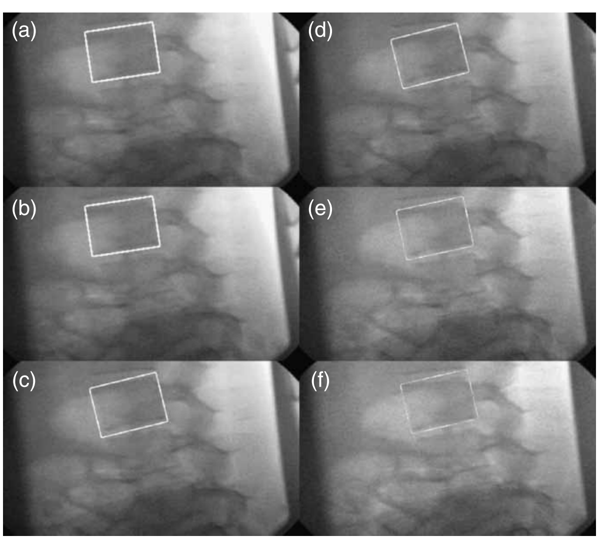


Figure 5. Example tracking results for the Male1 sequence. The frames are in alphabetical order. The white square indicates the estimated location and orientation of the tracked vertebrae. It can be observed that the estimated location and orientation is accurate across the entire video sequence.

Table 2. Computational time.

Processing time (s/frame)		
Penning et al. (2005)	Bifulco et al. (2001)	Proposed
84.92	95.49	47.25

are addressed during each step of the proposed algorithm in an attempt to achieve better tracking performance. Experimental results show that good tracking accuracy of lumbar spine vertebrae can be achieved for fluoroscopic image sequences when compared to existing methods. Future work includes an extensive investigation of different robust estimators to improve the number of correct feature point matches, as well as techniques for improving the pruning of mismatched vertebrae feature point pairs. Eventually the goal is to develop a system that accurately tracks lumbar spine vertebrae and provides useful motion information about the linear and angular displacements of the vertebrae in order to describe intersegmental motion. A robust, fast and easy-to-use vertebral tracking system is desirable for improving low back pain diagnosis, treatment and evaluating clinical outcomes.

References

Bifulco P, Cesarelli M, Allen R, Sansone M, Bracale M. 2001. Automatic recognition of vertebral landmarks in fluoroscopic sequences for analysis of intervertebral kinematics. Med Biol Eng Comput. 39:65–75.

Breen A, Allen R, Morris A. 1989. A digital videofluoroscopic technique for spine kinematics. J Med Eng Technol. 13: 109–113.

Bushberg J, Seibert J, Leidholdt E Jr, Boone J. 2002. Essential physics of medical imaging. Baltimore, MD: Williams & Wilkins.

Canny J. 1986. A computational approach to edge detection. IEEE Trans Pattern Recognit Machine Intell. 8(6):679–698.

Cholewicki J, McGill S, Wells R, Vernon H. 1991. Method for measuring vertebral kinematics from videofluoroscopy. Clin Biomech. 6:73–78.

Fauqueur J, Kingsbury N, Anderson R. 2006. Multiscale keypoint detection using the dual-tree complex wavelet transform. In: Proceedings of the IEEE conference on image processing.

Fischler M, Bolles R. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM. 24: 381–395.

Harada M, Abumi K, Ito M, Kaneda K. 2000. Cineradiographic motion analysis of normal lumbar spine during forward and backward flexion. Spine 25:1932–1937.

Harris C, Plessey M. 1988. A combined corner and edge detector. In: Proceedings of the 4th Alvey vision conference. p. 147–151.

Kalman R. 1960. A new approach to linear filtering and prediction problems. Trans ASME-J Basic Eng. 35–45.

Kovesi P. 2003. Phase congruency detects corners and edges. In: Proceedings of the Australian Pattern Recognition Society conference. p. 309–318.

Lowe D. 2004. Distinctive image features from scale-invariant keypoints. Int J Comput Vis. 60(2):91–110.

Muggleton J, Allen R. 1997. Automatic location of vertebrae in digitized videofluoroscopic images of the lumbar spine. Med Eng Phys. 19(1):77–89.

Okawa A, Shinomiya K, Komori H, Muneta T, Arai Y, Nakai O. 1998. Dynamic motion study of the whole lumbar spine by videofluoroscopy. Spine 23:1743–1749.

Penning L, Irwan R, Oudkerk M. 2005. Measurement of angular and linear segmental lumber spine flexion-extension motion by means of image registration. Eur Spine J. 14:163–170.

Schiele B, Crowley J. 1996. Object recognition using multi-dimensional receptive field histograms. In: Proceedings of the 4th European conference on computer vision. p. 610–619.

Schmid C, Mohr R. 1997. Local grayscale invariants for image retrieval. IEEE Trans Pattern Anal Machine Intell. 19(5): 530–534.

Selesnick I, Baraniuk R, Kingsbury N. 2005. The dual-tree complex wavelet transform. IEEE Signal Process Mag. 22(6):123–151.

Stewart C. 1999. Robust parameter estimation in computer vision. SIAM Rev. 41(3):513–537.

Takayanagi K, Takahashi K, Yamagata M, Moriya H, Kitahara H, Tamaki T. 2001. Using cineradiography for continuous dynamic-motion analysis of the lumbar spine. Spine 26: 1858–1865.

Teyhen D, Flynn T, Childs J, Kuklo T, Rosner M, Polly D, Abraham L. 2007. Fluoroscopic video to identify aberrant lumbar motion. Spine 32(7):E220–E229.

Wong A, Clausi D. 2007. ARRSI: automatic registration of remote sensing images. IEEE Trans Geosci Remote Sens. 45(5, Part II):1483–1493.

Wong S, Wong K, Wong W, Leong C, Luk D. 2004. Tracking lumbar vertebrae in digital videofluoroscopic video automatically. In: Proc. MIAR 2004, LNCS 3150. p. 154–162.

Wong W, Luk D, Leong C, Wong F, Wong K. 2006. Continuous dynamic spinal motion analysis. Spine 31(4):414–419.

Zheng Y, Nixon M, Allen R. 2004. Automated segmentation of lumbar vertebrae in digital videofluoroscopic images. IEEE Trans Med Imaging 23(1):45–52.

Downloaded by [University of Waterloo] at 12:43 11 July 2011