Analysing Video Sequences using the Spatio-temporal Volume

Toby Collins s0455374

MSc Informatics Research Review

November 2004

Introduction

In the overwhelming majority of studies to date, image sequences are primarily analysed and processed in groups of two frames, as by differentiating one frame from the other, one is able to infer the dynamics occurring in an image sequence. Although the two frame approach has been very successful in some applications, such as the MPEG compression standard, it faces considerable difficulties, if used for example to reason about non-constant velocity motion, the detection of occlusions and innovations, and long-term scene dynamics. It is also inherently related to the difficult problem of feature correspondence. This report aims to review the developments made in processing an alternative image sequence structure; the spatio-temporal volume, which has been research as a means to alleviate the shortcomings of the traditional pair-wise approach. The literature concerning the analysis of spatio-temporal volumes can generally be classed as either slice-based approaches, whereby two-dimensional temporal slices of the volume are processed, or volume-based, in which case three-dimensional structures are considered.

This division is reflected in the format of this report. An overview of the traditional motion analysis paradigm is first presented, and its use in two important tasks; Shape from Motion (SfM) and video segmentation is summarised. The next section provides a good introduction to the notion of spatio-temporal volumes. The next two sections reviews the recent work achieved in both spatio-temporal slice and volume analysis respectively. The final section summarises these findings, identifies the major shortcomings of current-state capabilities, and identifies possible directions for future development.

Motion for video sequence analysis

The analysis of motion enables us to extract visual information from the spatial and temporal changes occurring in an image sequence, and is a fundamental task in computer vision and image processing. Assuming illumination conditions remain constant, changes in an image sequence are caused by a relative motion between the camera and the scene; either by the viewing camera moving relative to a static scene, elements of the scene being in motion, or in the general case, both camera and objects moving independently. The problem of motion analysis may be divided into two sub-problems; that of feature correspondence and reconstruction. The correspondence problem concerns finding pairs of features in two or more perspective views of a scene such that each pair corresponds to the same scene point. Due to its inherent combinatorial complexity and ill-posed nature, feature correspondence is one of the hardest low-level image analysis tasks. The solubility of the correspondence problem is also influenced by factors such as image noise, periodic textures and object occlusion. The reconstruction problem states that, given a number of corresponding elements, and possibly knowledge of the camera's intrinsic parameters, what may be inferred about the 3D motion and structure of the observed world. The extraction of motion information from an image sequence has many important applications, with two of the most significant being the inference of Shape from Motion (SFM) and video segmentation.

Shape from motion

Shape from motion concerns recovering a scene’s 3D structure from motion-induced spatial and temporal changes occurring in an image sequence. With the knowledge of a camera's location, various perspective projections of a particular object allow us to infer its depth by comparing the projection's relative displacement in different frames. This is a hard problem for the similar reasons as stereoscopic vision; in particular, as it must solve correspondence problem. There is a vast range of approaches which address Shape from Motion, ranging from block matching algorithms to stochastic techniques, texture-based to feature based. A useful overview of such methods is provided by Jebera *et al.* [20].

Motion-based video segmentation

One of the primary goals of video analysis is to build a semantic interpretation of the scene being captured, which in itself involves the segmentation of the scene into its constituent semantic entities (e.g. objects or textures.) Although semantic-based segmentation, in which members of a scene are labelled with their real-world counterparts, operates at the most desirable object description level, it is a largely intractable problem, and in the general case is AI complete. Consequently, the majority of segmentation methods use concrete and measurable segmentation criteria that define non-semantic entities, and is typically motion-based or colour and texture. Motion-based segmentation relies on the fact that pixels associated with an object tend to move in a coherent fashion, which makes motion a very strong cue for object segmentation. Video segmentation has many important applications. These include video compression, in which it is possible to eliminate the redundancy related to the repetition of the same visual patterns in successive images. It is also used for video description tasks, such as logging, annotation and indexing. Automatic object extraction can help to enrich raw video content with object-specific information, which may be used by search engines and interactive multimedia documents. It is also useful in post-production, where special effects and visual modifications are applied to specific objects in the scene, and more generally in scene interpretation and video understanding.

Current approaches in motion analysis

The fundamental goal of motion analysis is to determine a vector field describing changes in the image over time [18]. The most widely researched techniques for doing so can be broadly separated into three groups; *gradient-based*, *correlation-based* and *feature-based* approaches. Gradient-based methods make use of spatio-temporal partial derivatives to estimate the image flow at each point in the image. Horn and Schunk [18] used the spatio-temporal derivatives of the image brightness function, which assumes that the brightness of any part of the imaged world varies very slowly, so that the derivative of the brightness is zero. Gradient-based motion segmentation is often performed by first recovering a dense optical flow field and then fitting this field to a model, which is often affine [5, 19, 23, 58]. Since reliable computation of optic flow often requires expensive computations, these methods are mostly limited to off-line applications. A number of methods have been developed for simultaneously recovering motion and performing segmentation [9, 40, 46]. In these techniques, segmentation is reformulated as a Markov Random Field (MRF) based relaxation problem. Correlation-based techniques determine the motion vectors by comparing the similarity in intensity patterns between two images in the sequence [32]. This method is generally used to aid the matching of image features or to find image motion once features have been determined by alternative methods.

Feature-based approaches aim to compute and analysing the optic flow at a small number of well-defined image features in a scene, such as corners, edges, blobs. In essence, this method operates in a feature tracking framework, where each frame in the sequence is first spatially segmented, and the extracted features are matched with those corresponding features in later frames. The simplest and most popular approach involves two consecutive frames, from which two sets of features are extracted; whose matching gives rise to a single set of motion vectors. Another feature-based method involves using the features in one frame as seed points, and then using other methods, such as gradient-based, for flow detection [52].

Analysis using the spatio-temporal volume

The approaches mentioned above each share the same characteristic of typically determining motion based on two frames in the image sequence, and so share similar shortcomings as mentioned earlier. Motion estimation using frame differencing is also highly sensitive to noise, and results in a high false positive rate which is hard to surprises. Gradient and feature-based approaches also share the characteristic that they each favour features in either the temporal or spatial domain respectively; the first finds temporal features (e.g. optical flow), and then groups these spatially, and the second first finds spatial features, and projects these temporally.

A more recent approach is to unify the analysis of spatial and temporal information, by constructing a volume of spatio-temporal data in which consecutive images are stacked to form a third, temporal dimension (figure 1). The benefits of analysing this volume are realised when the images are sampled sufficiently often such that there is continuity in both temporal and spatial domains. One of the major advantages of this representation is that by analysing feature structures in this volume, we may reason about much longer-term dynamics. Also, by conjointly providing spatial and temporal continuity, the complexity of feature correspondence is significantly reduced. A further advantage is that occlusion events are made much easier to detect, as they are represented explicitly in this volume as truncated paths [26, 57].

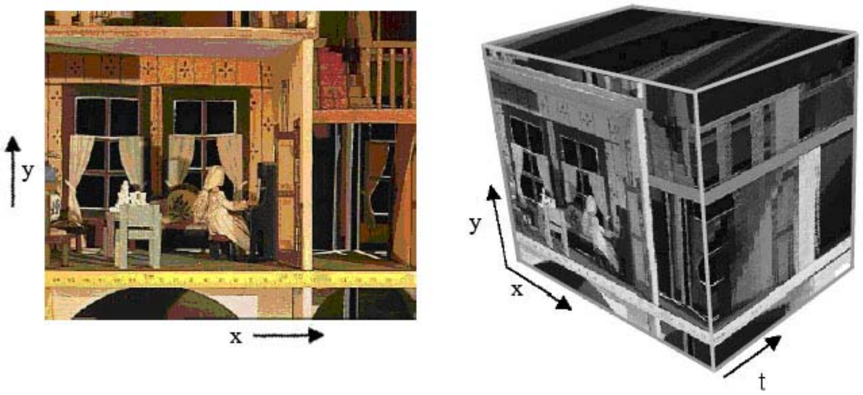


Figure 1 An example spatio-temporal image volume. On the left is the first image of the scene and on the right is the video cube constructed by zooming into the doll's house

The spatio-temporal volume was first pioneered in 1985 by Aldelson and Bergen [1], in which motion models were based on energy and impulse response to filters. Since then, the spatio-temporal volume has been predominantly studied in image processing, both as a means for inferring a static scene's depth information, and for performing segmentation of dynamic scenes. For both of these tasks, the methods developed can be grouped depending on whether the volume is processed by analysing 2D structures found in temporal slices, as a whole to analyse paths and surfaces generated in these volumes caused by relative camera/object motion. The following two sections summarises the work conducted for in both of these approaches.

Pattern analysis of spatiotemporal slices

One way to analyse the spatio-temporal volume is to consider it as being formed by a stack of two-dimensional temporal slices. For example, if the cube in figure 1 were to be sliced horizontally, one slice per scan line, then each slice exhibit structures related to the image features which pass over that scan line over time. These slices have been studied in for a variety of problem domains: to infer feature depth information [6, 8] , generating dense displacement fields [31, 51], camera work analysis [28, 42, 44], motion categorisation [41, 43], the detection of parked vehicles [17], ego-motion estimation [48], for use in advanced navigation systems [22], view synthesis [53] and gait recognition [45]. A selection of the key techniques for this approach is presented below.

Epipolar plane image analysis

Slices of the spatio-temporal volume were first investigated by Bolles *et al.* [8], which focused on the geometric recovery of static scene structure. The particular class of slices analysed were termed *epipolar plane images* (EPIs), and by restricting camera motion to linear paths, with a fixed orientation orthogonal to the direction of motion, depth information could be extracted from the relative angles of paths formed by features in the EPI. The concept of EPI analysis can be best explained diagrammatically, and the general framework is illustrated in (figure 2).

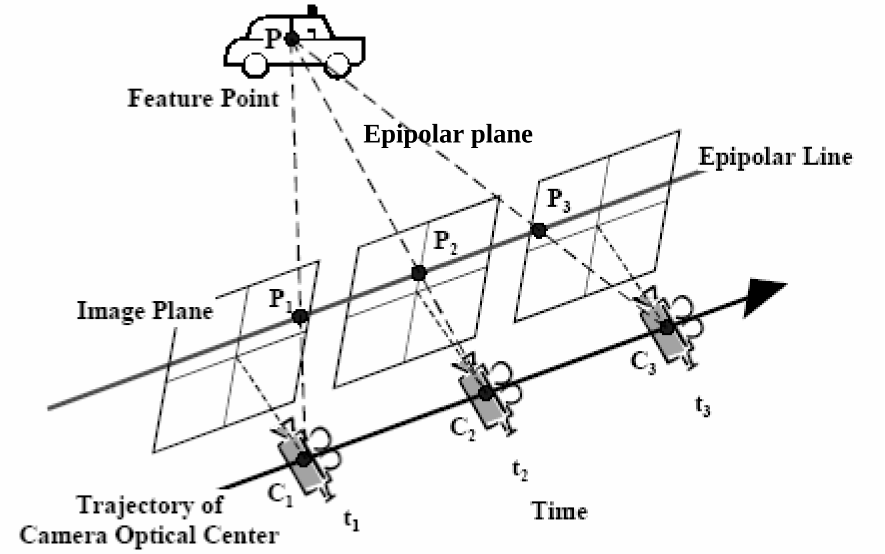


Figure 2 The framework for Epipolar Image Analysis

For any feature point P , we first define an *epipole plane* to be the plane passing through P and any two camera positions. This plane is identical for any pair of collinear camera positions. We further define the *epipolar line* to be the intersection between an epipolar plane and any of the cameras' image planes. Given the camera's restriction to linear motion with a fixed, orthogonal viewing angle, each epipolar line passes horizontally through the image planes, and occurs with the same vertical position. With these constraints, we may define an EPI for a given epipolar plane to be a slice in the spatio-temporal volume which passes through each camera's epipolar lines. This corresponds to a horizontal slice in the temporal domain, which is illustrated in figure 3.

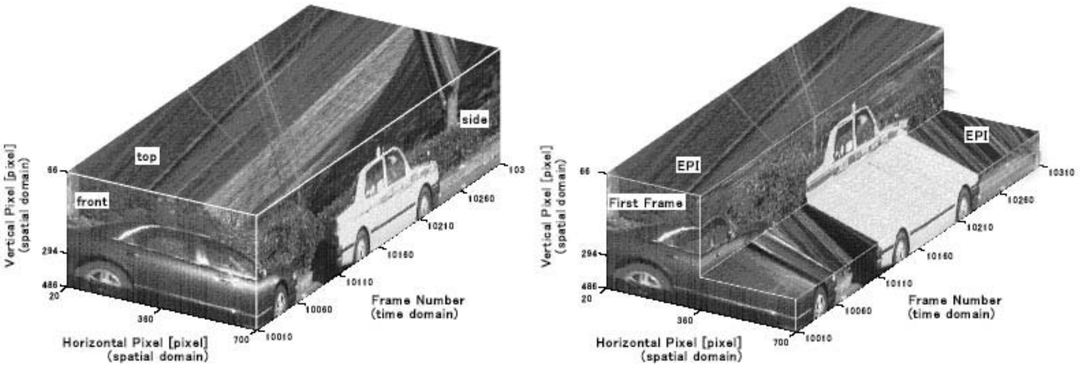
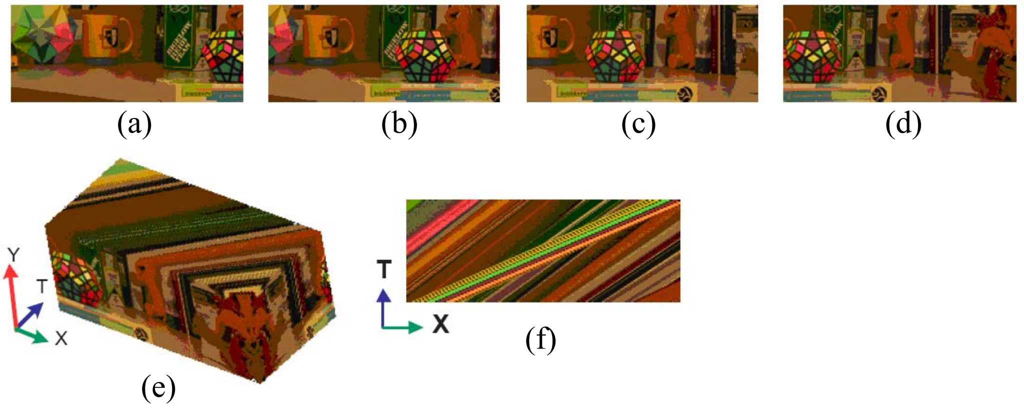

top
side
front
Vertical Pixel [pixel] (spatial domain)
66
294
486
20
360
700
Horizontal Pixel [pixel] (spatial domain)
103
10260
10210
10160
10110
10060
10010
Frame Number (time domain)
EPI
EPI
First Frame
Vertical Pixel [pixel] (spatial domain)
66
294
486
20
360
700
Horizontal Pixel [pixel] (spatial domain)
10310
10260
10210
10160
10110
10060
10010
Frame Number (time domain)

Figure 3 Illustration of Epipolar Plane Images. The spatio-temporal volume is given by the image on the left, which has been created by a lateral motion of a camera mounted on a moving vehicle. The second image shows an example EPI.

The advantage of partitioning motion analysis along EPIs is that any given feature will reside on a single EPI throughout the spatio-temporal volume. Consequently, the problem of stereo correspondence has bee reduced from two dimensions to one, as it is only necessary to search along the corresponding epipolar line in the other image. Furthermore, the strictly linear camera motion causes features to trace straight paths through the EPIs (figure 4), which can be easily extracted (Bolles *et al*. used Ramer's algorithm to fit line segments to the zero crossings of the slice convolved with the Laplacian edge detector.) The primary advantage of EPIs analysis is that it essentially combines the acquisition and tracking stages of conventional motion analysis into one.


(a) (b) (c) (d)

(e) (f)

Figure 4 Example linear features found in EPIs. Images (a),(b),(c) and (d) shows selected frames from a video sequence taken with a camera moving from left to right of a static scene. Image (e) shows the spatio-temporal volume. Image (f) shows an EPI, illustrating the linear path features and truncations caused by the coloured shape occluding features behind it.

Extracting Depth from EPIs

Figures 3 and 4 illustrate how lateral camera motion causes features in space to trace continuous straight paths in EPIs. The slope of these paths is related to the distance the feature is to the camera's centre. This relationship is illustrated in figure 5, which shows the change of parallax when a camera moves between two locations, C_1 and C_2 for a feature P .

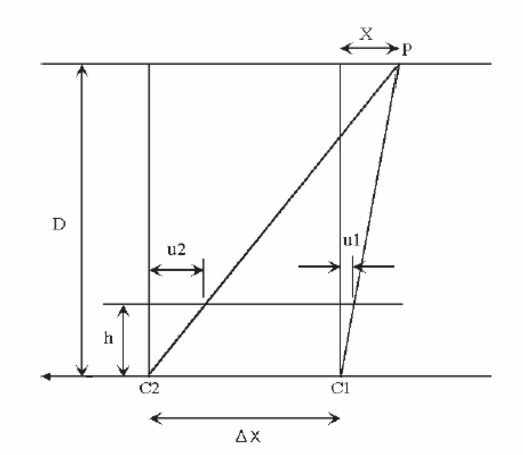


Figure 5 The geometric relationship between feature depth and the angle the feature is projected onto two camera positions.

The depth and the parallax are related by the following expression:

$$\Delta U = u_2 - u_1 = \frac{h(\Delta X + X)}{D} - \frac{hX}{D} = \Delta X \frac{h}{D}$$

∇U may be found using the corresponding points of P in the EPI, and ∇x is a function of camera speed. Knowing these parameters, an estimate for the depth of the feature can be easily obtained.

Shortcomings of EPI analysis

Despite the considerable advantages of EPI analysis, the method of Bolles *et al.* suffered form the major constraint of linear camera movement, which was necessary to ensure that each 3D feature remains on the same EPI throughout the image sequence. Also, features may only trace linear paths in the EPI if the camera's orientation is fixed and orthogonal to the direction of motion. A further problem is that the approach also does not account for independently moving objects.

Generalising EPI analysis

In light of these limitations, a succession of developments on EPI analysis has aimed to alleviate these restrictions. The first was made by Baker *et al.* [6] and relaxed the constraint for orthogonal and fixed camera motion to arbitrary, known orientations. Arbitrary camera orientations causes hyperbolic feature trajectories in the spatio-temporal volume, but by mapping these points in the spatio-temporal volume (x, y, t) to an epipolar cylindrical coordinate system (r, h, θ), where θ is the epipolar-plane angle for a particular view, the trajectories were once again made linear (figure 6.)

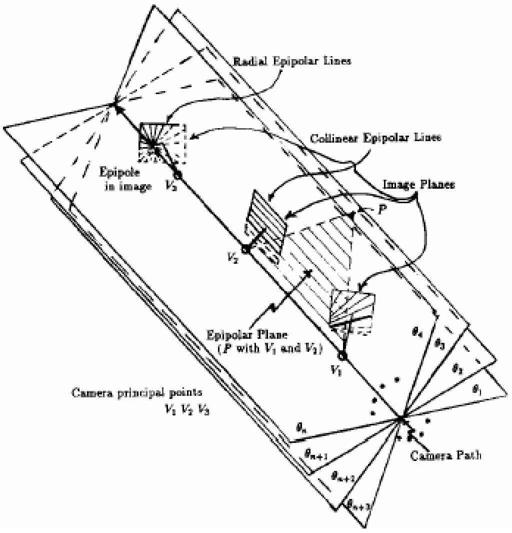


Figure 6 EPI analysis for linear camera motion in which the camera orientation may be arbitrary. V_2 shows a camera whose orientation is orthogonal to the direction of motion. V_1 and V_2 are at arbitrarily orientated. The intersection of each epipolar plane radiates out from the camera path, and by mapping each frame into a cylindrical system, point P will trace a linear path through this new spatio-temporal volume.

A more recent development to allow for non-linear camera motion was made by Li *et al.* [31], which introduced the notion of piecewise linear EPI analysis. In this framework, EPIs are constructed only from those images of the sequence where the assumption of linear equidistant camera motion is approximately fulfilled. Lie *et al.* investigated a camera rotating on a predefined circle, in which small arc segments may by approximated by straight lines, which are then utilised to determine depth of corresponding points. Unfortunately, this approach significantly reduces the amount of reference images available for 3D reconstruction.

In 2003, Feldmann *et al.* [15] extended EPI analysis to accommodate other parameterised camera movements. Specifically, path structures in the spatio-temporal cube, termed *Image Cube Trajectories* (ICTs) resulting from concentric circular movement for orthographic and perspective cameras were shown to be very well defined. Path detection algorithms were therefore adapted to detect those paths expected in the volume. In contrast to standard EPI analysis, an ICT is constructed in reverse. Firstly, the ICT for a particular feature is determined by its image position and assumed depth. In a second step, the image cube is used to test such a path exists. If not, the depth is changed until the resulting ICT fits to the image cube. This approach extends to define rules only for other parameterized camera movements, such as parabolic camera paths.

Epipolar Plane Image analysis applications

A recent example of the use of EPI image analysis has been for image understanding for street parking vehicle detection [17]. A side-facing camera is mounted on the side of a moving vehicle and images are captured at a constant frequency at a height that cut the moving cars. Feature paths were detected using the Canny edge detector, and the Hough transformation was used to detect any concealed lines. The strongest peak of this transform is selected to base the distance measure. Their method achieved a detection rate of 76.9%.

EPI analysis has also been used for automatic texture image database construction, to enhance navigation systems with real images [22]. Kawasaki *et al.*'s key concept was to introduce the idea of EPI-EPI matching. Through the use of models of objects such as buildings from digital maps, they proposed that virtual EPI models could easily be generated by simulating the camera motion and parameters. Given these simulated EPIs, their idea was to match these with the real EPIs from the video data, to relate the digital map with the video data. The matching algorithm used was based on DP matching. The major advantages of this approach are that lines themselves do not have to be detected on real EPIs as accurately as with usual EPI analysis, and that the camera track does not have to move in a straight line.

Spatio-temporal slice processing for camera work analysis

Another area of research which has developed the use of temporal slice analysis has been in automated camera work analysis. Video footage typically consists of a series of shots, where each shot is an uninterrupted sequence of frames from a stationary or moving camera, and robust scene change detection is an important component of content-based video browsing and video summarisation. The boundary between shots is typically demarcated by cuts; an instantaneous shot transition, wipes; a cross-faded gradual transition, and dissolves; the gradual fading between shots. The majority of algorithms for detecting scene change may be categorised as statistic-based, feature-based and motion-based and predominantly operate using a frame-to-frame similarity measure.

Ngo *et al* [44] first used spatio-temporal slices for the detection of cuts and wipes, where the task of detecting scene breaks was reformulated as the detection of boundaries in spatio-temporal slices. Therefore, the problem of video segmentation has been reduced to a problem of image segmentation. The approach used the analysis of two orthogonal slices, one horizontal and one vertical, taken through the centre of the spatio-temporal volume (figure 7 and figure 8).

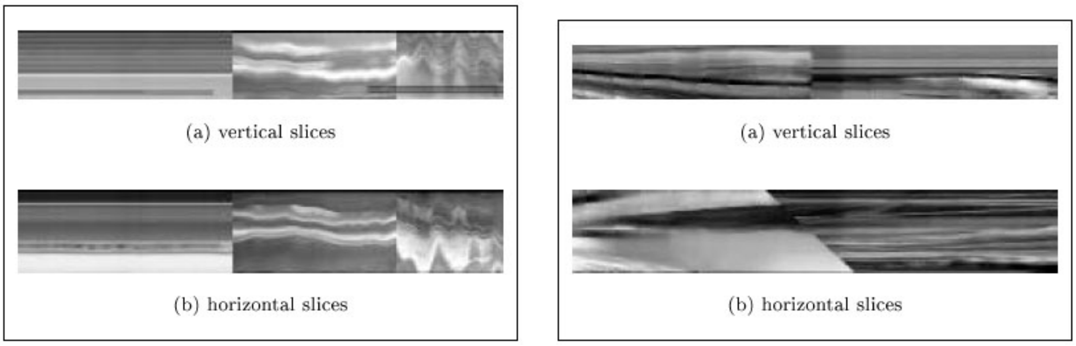


Figure 7 Boundary characteristics in horizontal and vertical spatio-temporal slices for camera cuts (left) and wipes (right), separating similarly textured scenes belonging to one scene.

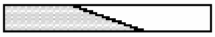
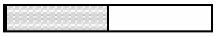
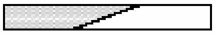
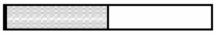
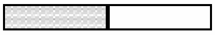
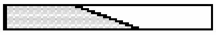
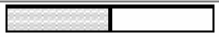
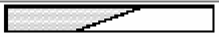
wiping direction	horizontal slices	vertical slices
left-to-right		
right-to-left		
top-to-bottom		
bottom-to-top		

Figure 8 Boundary orientations for different styles of wipes

The slices were analysed by first convolving with the first derivative Gaussian, and then processed using Gabor decomposition, in which the real components of multiple spatial-frequency channel envelopes are used to from a texture feature vector. A Markov energy-based image segmentation algorithm is then used to locate the colour-texture and classify discontinuities at region boundaries. Evaluated on news sequences, documentary films and movies, their approach performed at approximately 95% for cut detection, but only 64% for wipe detection.

Ngo *et al*. [43] further extended their work to incorporate gradual transitions caused by dissolves, which usually defeat traditional statistical analysis techniques. They also improved the robustness of their system by incorporating a diagonal spatio-temporal slice, and achieved wipe detection accuracy of only 64%.

Analysing gait using spatio-temporal slices

Niyogi *et al.* [45] were the first to suggest human gait could be analysed using the special signatures generated by walking in space-time. Two slice regions crucial to their method are shown in Figure 9 and Figure 10.

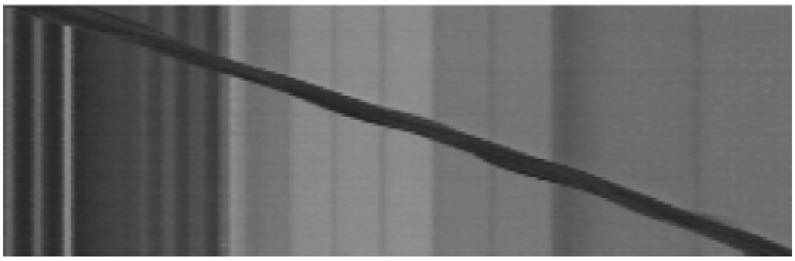


Figure 9 An XT-slice taken at the walker's head height, indicating the head mostly only undergoes translational movement during walking.

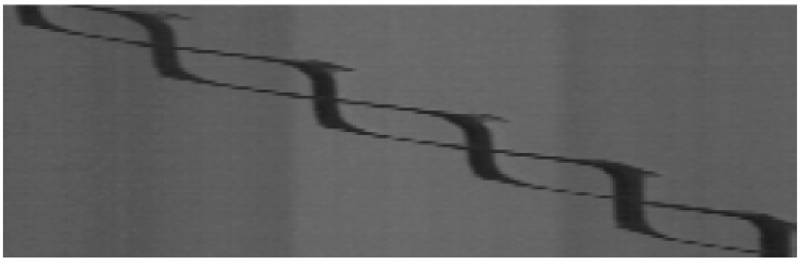


Figure 10 A slice taken at the height of the walker's ankles. The criss-crossing of the walker's legs as the walker moves from left to right is given as a unique braded signature for walking patterns

Niyogi *et al.* process these patterns in a four-stage recognition architecture. Gait is first detected by finding translating objects in an image sequence and testing whether they contain the braded pattern (figure 10) in the lower half of the translating object, which corresponding to ankle motion in XT. The translating objects were found using a simple change detection algorithm between each image and the background, and the test for the braded pattern was performed using a best template match over a small number of amplitudes, periods and skew which parameterise the braded pattern.

Once gait is detected, the rough estimate of the walker's pattern is refined using Snakes [21]. Snakes are active contour models using "an energy minimizing spline guided by external constraint forces and influenced by image forces that pull it toward features such as lines and edges" [21]. Given an initial list of points that define the snake, the snake will 'climb' to the local maxima in the energy function. The energy function used by Niyogi *et al.* is the slice with maximal correlation with the braded templates.

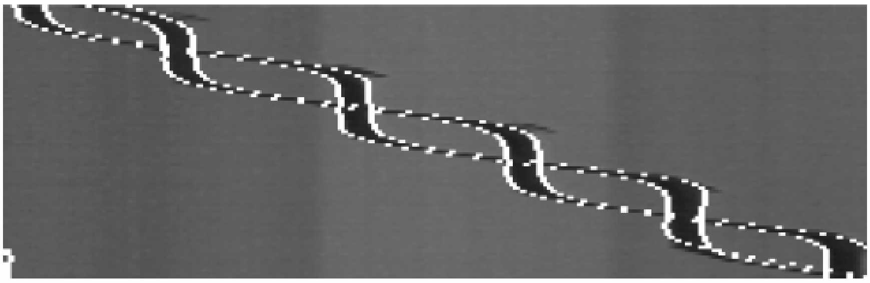


Figure 11 4 snakes are used to model the braded signature, ehich are attracted to the positive and negative spatial derivatives of the braded pattern.

The entire body is then modelled using such snakes; for each slice from head to toe. Near the hip, the two snakes ideally merge into one. Gate is then modelled by averaging the body contours to form two 'skeletons', and the location of the head, hip, knee and ankle joints are estimated using another snake operation for all XY slices of the image sequence (Figure 12). As there are second order discontinuities at these joint locations, the Snakes are set to be second order discontinuous at coarse locations for a simple height model of a human. The angle signals at these joints, which vary as a function of time, are extracted from a stick model parameterised by these joint locations, and are then classified using a table of previously observed gait signatures, using a standard k-nearest neighbour classifier.

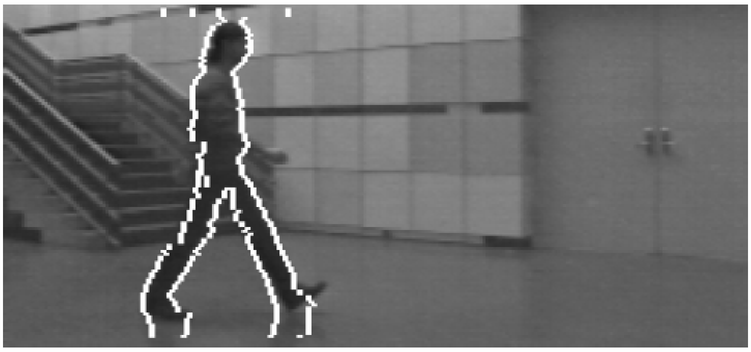


Figure 12 an example frame from the walking image sequence, with the four fitted snakes overlaid in white

The algorithm was run on 24 different image sequences, and performed at a recognition rate of 79%. In their approach, the camera is fixed, the walker walks at mostly a constant speed, the direction of walk is roughly lateral relative to the camera, and no obstacles carried by the walker are present.

Motion as orientation in the spatio-temporal volume

In addition to the motion analysis techniques utilising spatio-temporal slices, there have been several approaches to perform the same function based on analysing the entire spatio-temporal volume. Otsuka *et al.* [49] propose a new framework based on image motion trajectories in spatiotemporal space for a static camera recording dynamic scenes. *Trajectory surfaces* were formed by edges and contours of images in spatio-temporal space using Hough transforms. These trajectory surfaces were then used as a means to estimate the velocity components of the objects in a scene, which were determined by the orientation of the intersection line formed by tangent planes to the trajectory surfaces. One of the shortcomings of this work is that the camera motion must be parametric. If more complex camera motion is considered, such as piecewise rotation and translation, not only is the creation of parameterised motion models not feasible, but with a parameter space of more than two dimensions, the computation of the Hough transform becomes unfeasible.

In contrast to the work of Otsuka *et al.*, Rodrigues *et al.* [54] have recently proposed a SfM method which exploits curves in the spatiotemporal volume, by assuming known camera parameters, but accounts for arbitrary (including non-smooth) motion parameters. It also assumes camera parameters such as focal length, trajectory and orientation are well estimated. Their goal was to solve which 'interesting' 3D points generated a set of implicit curves found in the spatio-temporal volume. The 'interesting' points were those which lay on contours of the image frames. Depth estimates were established by finding a minimum match error between the implicit spatiotemporal curves with a set of candidate depth curves for a particular interesting feature P .These candidate depth curves were generated by reverse-projecting P with known camera parameters and an attributed depth estimate.

The performance of this method was evaluated both on synthetic and real scenes. The synthetic scene comprised 20 coloured boxes, arranged in a circle (figure 13), used to demonstrate the method on a high number of occlusion occurrences. Promising results can be seen in the reconstructed seen, as shown in figure 10, and the visible artefacts are mostly due to the reduced number of interesting points used.

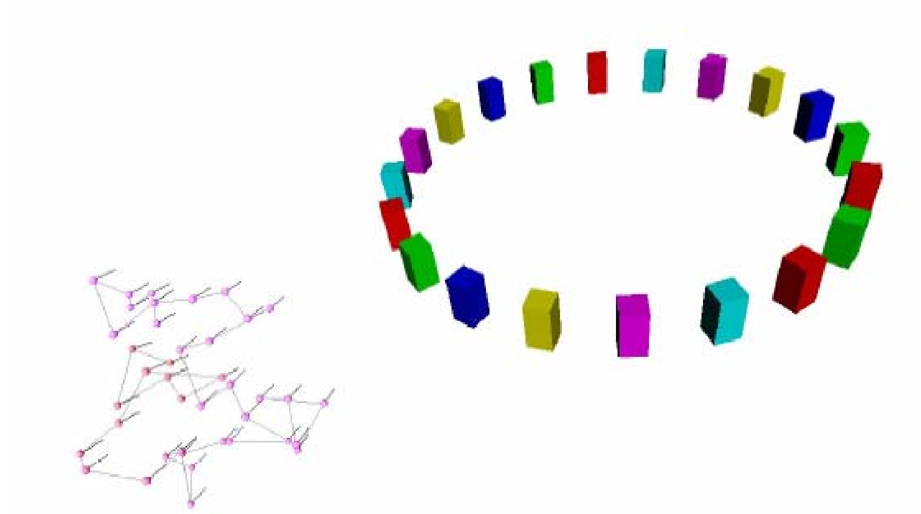


Figure 13 The collection of coloured boxes is shown to the right, while the small line segments extending from each dot represent the camera's orientation. The line connecting them is the camera's path

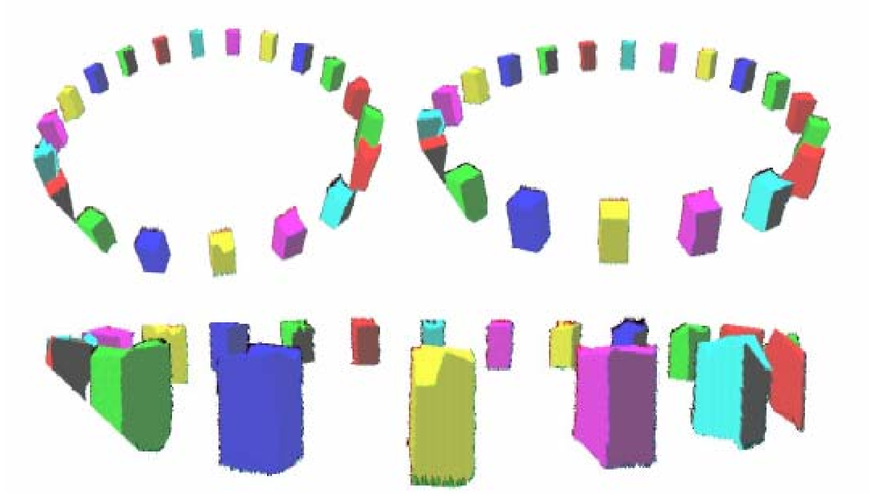


Figure 14 Three views of the reconstructed scene

The real scene comprised three shapes in a plane background. Three of the 40 scenes shot are shown in figure 15. Figure 16 shows four views of the reconstructed scene.

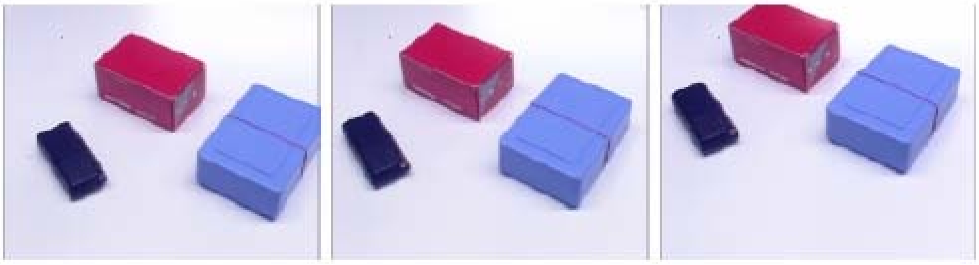


Figure 15 Three examples of the real scene image sequence

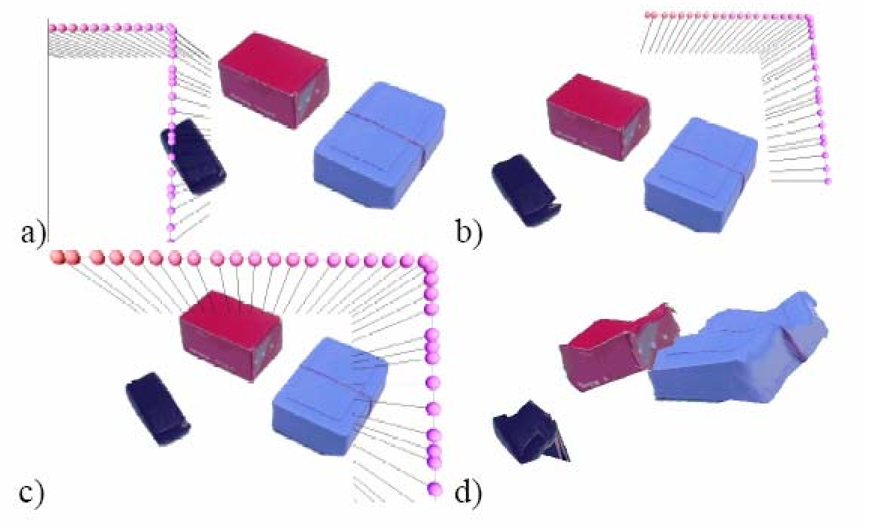


Figure 16 The reconstructed scene taken from 4 views

3D segmentation using the spatio-temporal volume

The alternative approach using the spatio-temporal volume for scene segmentation has been to process the structures found in the entire volume, rather than by analysing distinct slices. A variety of techniques for doing so have been studies. Of these include; spatio-temporal manifolds [6, 7], the 3D structure tensor [29, 41, 43], mean shift analysis [13], Fourier analysis [47] and deformable shape models [17]. Surprisingly little work has been done on fitting spatio-temporal surfaces with active surfaces [59]. Very recently, level set evolution equations have been successfully used for spatio-temporal segmentation [24, 25, 26, 27, 55, 56], [14, 33, 34, 35, 36, 39]. A selected overview of these techniques is now presented.

Motion segmentation using 3D structure tensors

The spatio-temporal volume has been analysed using 3D structure tensor-based optical flow [29, 41, 43], which exploits the orientations of local grey value structures within the frame stack. Moving and static parts of the image plane can be determined by the direction of minimal grey value change in the spatio-temporal volume. The advantage of integrating spatial information with tensor-based optical flow is that it allows for more reliable motion calculation which suppresses background noise by considering multiple frames in the video sequence (Figure 17.) The number of such frames is termed the support window of the structure tensor. For example, [41, 43] use a support window of 3^3, meaning that the motion calculation for each pixel is performed within a spatio-temporal area of $3 \times 3 \times 3$ pixels, while Kühne *et al.* [29] developed a coarse-to-fine hierarchy of support windows.

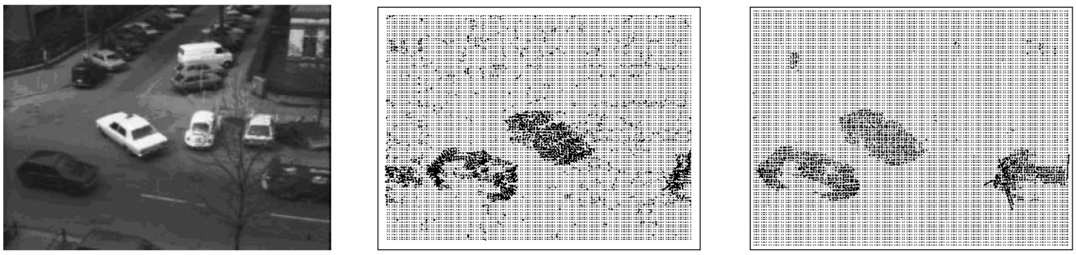


Figure 17 The image on the left is frame 10 from the well known Hamburg Taxi sequence. The middle image shows the optical flow computed with the Lucas Kanade algorithm. The right

image shows the optical flow computed with the 3D structure tensor. Here, the background noise is eliminated without any pre-filter, and even small structures like the pedestrians have been identified.

For an image sequence $I(X)$, where $X = (x, y, t)^T$ and x and y are the spatial components and t is the temporal component, the 3D structure tensor is given by $J(x, y, t) = h(x, y, t) * (\nabla I \cdot \nabla I^T)$, where $\nabla = (\partial_x, \partial_y, \partial_t)$ denotes the spatial and temporal gradients, and $h(x, y, t)$ is the spatio-temporal filter for a given support window.

Purely tensor-based segmentation was found not to be effective in accurately distinguishing the boundaries of objects in a scene alone. Firstly, areas of constant grey within the moving objects do not receive dense motion vector fields. Secondly, the tensor fails to provide the true object boundaries accurately since the calculations within the neighbourhood blurs motion information across spatial edges. 3D structure Segmentation methods based on 3D structure tensors have been developed which can be further classified as either contour or region based. Contour-based segmentation aims to refine the contour models based on the motion masks generated from the motion field. A tensor-based optical flow field is used by Kühne *et al.* as the external forces to converge a geodesic active contour model in addition to the boundaries of the moving object. Geodesic active contours were used to group neighbouring regions and close holes and gaps, are topological flexibility and allow the simultaneous detection of multiple objects.

The approach of Kühne *et al.* so far experiences problems when considering sequences containing large velocities, as if the displacement exceeds the size of the local neighbourhood, the motion of the feature cannot be detected. To overcome this, a hierarchical algorithm was developed which embeds the structure tensor technique in a linear scale-space framework. Consequently, the calculations are performed in a coarse to fine manner, using a Gaussian pyramid, whereby motion vectors determined at coarser levels in the pyramid serve as an initial guess for subsequent refinement levels, until the highest resolution is reached.

The segmentation algorithm was applied to two real-world sequences, the first one is the Hamburg Taxi sequence, and good results can be seen in Figure 18.

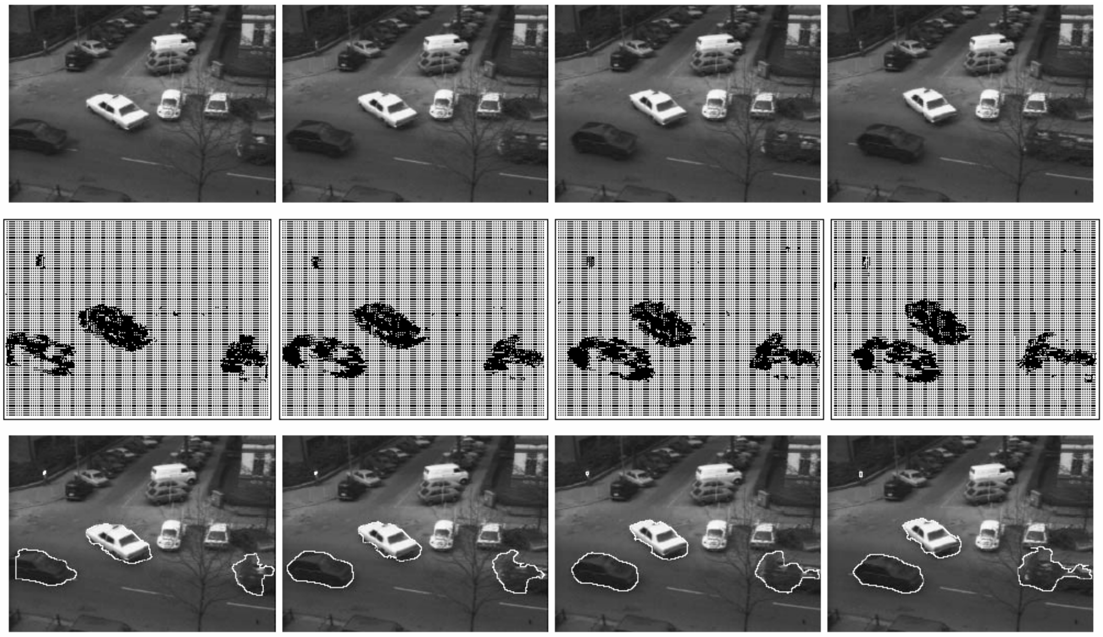
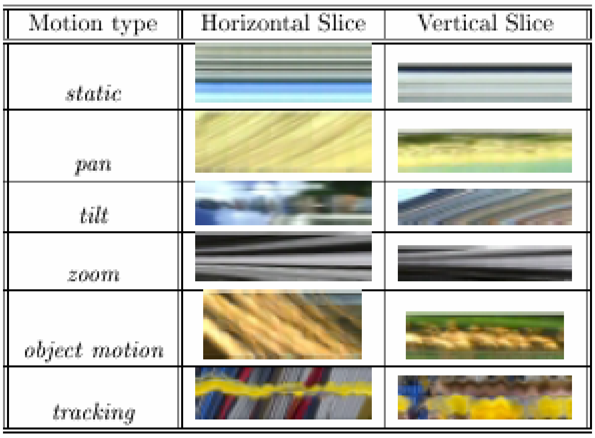


Figure 18 Results of [29] for the Hamburg taxi sequence. The first row of images are samples from the video sequence. The middle row is the results of the 3D structure tensor algorithm. The final row shows the segmentation using geodesic active contours.

Categorising camera motion using 3D Structure Tensors

3D structure tensors were also used in [41] and [43] to categorise dominant camera motion (static, pan, tilt, zoom, and tracking), in addition to object motion. Figure 19 illustrates example 2D slice patterns associated with these types of motion.



Motion type	Horizontal Slice	Vertical Slice
static		
pan		
tilt		
zoom		
object motion		
tracking		

Figure 19 Temporal slice patterns for various types of camera motion. Note that for tracking motion, the slices exhibit both panning and static elements.

By investigating the distributions of motion orientations of all temporal slices, motion types can be classified, as can different motion layers. In [43], this distribution is approximated using 3D structure tensors to form a 2D tensor histogram, $M(\phi, t)$, who's dimensions are a 1-dimensional orientation histogram for each temporal slice, and time. The value at each point in the histogram is given by a degree of confidence which each pixel has for a particular orientation. Dominant motion trajectories could then be traced by tracking the peak histogram values over time (Figure 20.)

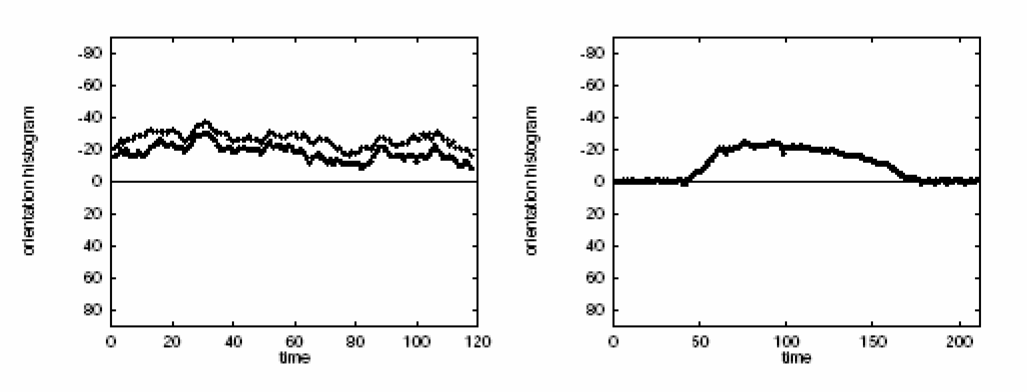


Figure 20 Two example motion trajectories based on tensor histograms. On the left is the histogram peaks for two parallax panning shots, and on the right is a camera moving from static, to pan, to static again.

Experiments were conducted on two standard test videos, *lgerca_lisa_1,mpg* and *lgerca_lisa_2.mpg*. On average, dominant motion categorisation was performed at 12 frames per second. The algorithm was extended in [41] to account for finer-grained motion categorisation. This was performed by two techniques which traded off simplicity with effectiveness. The first partitioned the spatio-temporal volume into sub volumes by employing k-mean clustering to group similarly coloured temporal slices. Ideally, each sub-volume corresponds to the evolution of one moving object over time. However, this fails when the background, for example, is composed of various colour elements. The second exploited the motion trajectories inherently existing in the tensor histograms. For tensor histograms containing multiple motion trajectories, the idea is to simply back-project these trajectories to the spatio-temporal slices to form spatially-separated motion layers.

Fitting deformable models to spatio-temporal surfaces

Deformable models

Although, much work has been done on the tracking of rigid objects in 2D sequences, the structures formed in the spatio-temporal volume are inherently non-rigid. One popular approach for modelling non-rigid, time varying objects is through the use of deformable models. One such example is the use of Snakes [21] and their variants [11], [37], and are used widely for segmenting non-ridid objects in 2D and 3D (volume) images. However, there are several well-known problems with Snakes. They were originally designed as interactive models, and so rely upon a user overcoming initialisation sensitivity. They were also designed to be a general model showing no preference for a particular object shape other than those that are smooth. Consequently, Snakes do not perform well in the face of shape abnormalities caused by occlusion, irrelevant structures or noise. In response to these deficiencies, techniques were developed which incorporated a priori knowledge of object shape, the most predominant being Active Shape Models [12], whereby the statistical variation of shapes is modelled using a set of training examples to fit an example of the object in a new image. The shapes are constrained by a Statistical Shape Model to vary only in ways seen in a training set of labelled examples. Dynamic deformable models [30] were later developed which described the shape changes over time in a single model which evolves to reach a state of equilibrium where internal forces, representing constraints on shape smoothness, balances the external image forces.

Recently, Hamameh *et al.* [16] have extended 2D Active Shape Models to deformable spatio-temporal shape models. Similar to 2D ASMs, a single static shape is represented by a set of labels or landmarks, $\{x_i(t), y_i(t)\}$,where x and y are landmark coordinates, i is the landmark number and t denotes time. The segmentation technique they developed is based on deforming a spatio-temporal shape to better fit the image sequence data only in ways that is consistent with the training set. To segment a similar time-varying object in a new image sequence, we start with an initial ST shape model (e.g. the mean ST shape) and an initial pose estimate. This ST shape model is then deformed by minimising an energy function using dynamic programming, and repeated until the energy function converges.

The method was tested using only synthetically generated data, with added synthetic noise and mild occlusions. Both the x and y coordinated for the generated ST shape moved in accordance with a sinusoidal function with certain amplitudes and frequencies (Figure 21.) An example of the model fitting to a sequence of test images is shown in Figure 22.

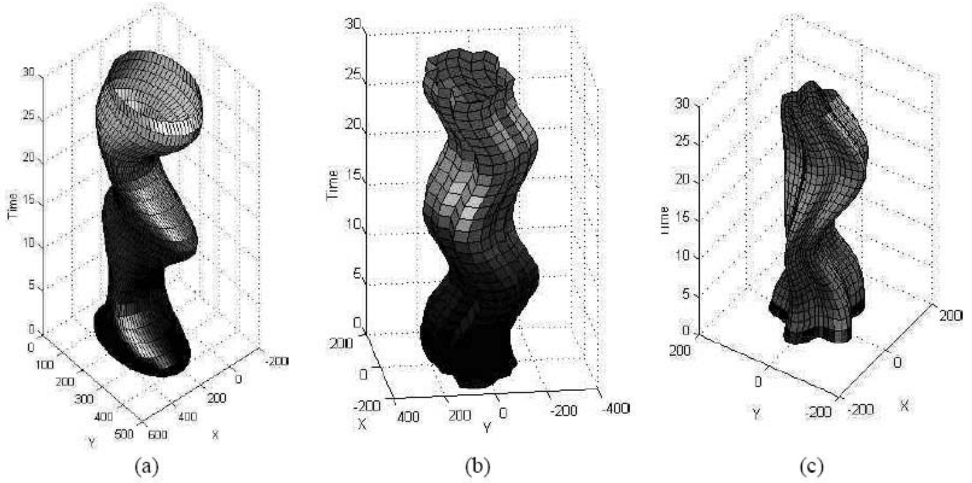


Figure 21 Examples of synthetic spatio-temporal shapes. (a) shows a circle with translational motion, expanding and shrinking in time, (b) is a 'random star' with translational motion. (c) is a 'Sinusoidal star' with translational motion, whilst both expanding and shrinking in time.

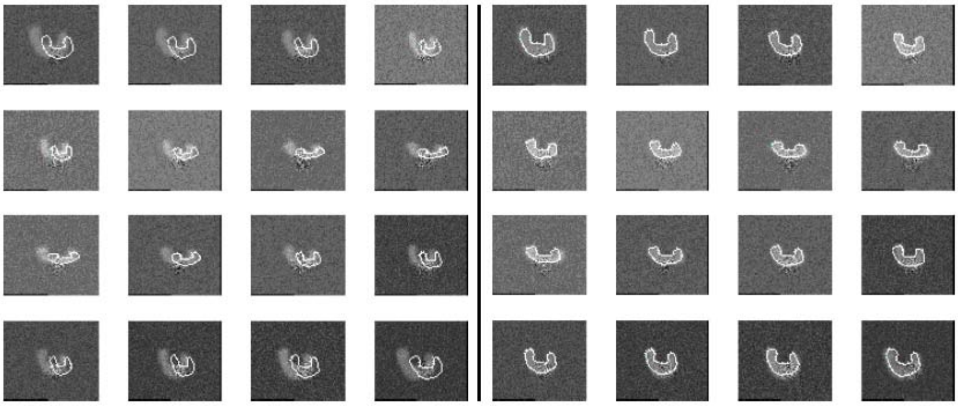


Figure 22 Spatio-temporal segmentation results, with strong local noise and moderate global noise in all frames. Frames 1-18 are presented on the left, and frames 19-32 are on the right. The deformable spatio-temporal model for each frame is shown in white, and after 32 iterations, fits the test data extremely well

Hamameh et al.'s work concluded by remarking they are considering a multi-resolution extension and a time-scaling and time-translation feature.

Level set methodologies

A very recent and active development in spatio-temporal segmentation has been the parameterising of surfaces in the spatio-temporal volume using level set methodologies. Similar for the active surface approach, the unknown surface to be estimated is parameterised as an active surface, but rather than solving using an active surface approach, the resulting cost functional is minimised using the level set methodology. By embedding the surface into this higher-dimensional function, the problems of the original active-contour formulation concerning stability and fixed topology are alleviated. The methods do not require a known background or require estimation of the image motion field. Furthermore, optical velocities can be estimated along motion boundaries from geometrical properties of the spatio-temporal surface [14]. Since 2002, there have been two independent research groups involved in developing this approach; the first at Boston University, and the second at the INRS-Tellicommunications in Quebec, Canada. Both groups formulate this problem in the framework of maximum *a posteriori* probability (MAP) estimation.

Konrad *et al.* [24] at Boston University first studied this approach in a 'volume completion' framework. Their models were relatively simple, and did not permit moving backgrounds due to camera motion, or multiple objects. Also, the computational complexity of this approach was reported to be significant. However, their approach showed excellent object shape recovery and tracking between frames (Figure 23.)

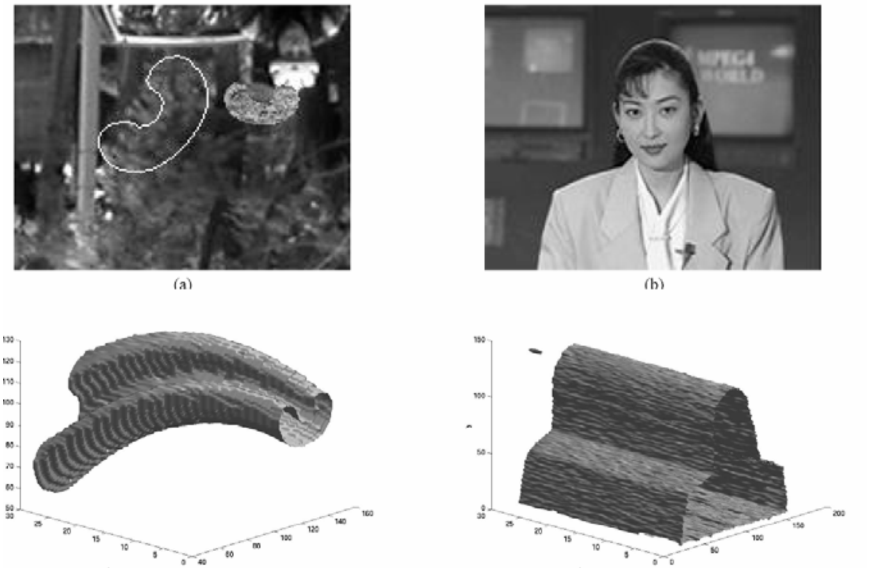


Figure 23 Object shape recognition and tracking performed in [24] computed jointly over 30 frames. In the case of the bean (left), the algorithm performs admirably in the face of significant motion, and zoom-in.

Konrad *et al.* further extended their work by demonstrating that the 'object tunnels' formend in the spatio-temporal volume could be effectively used for the detection and characterisation of occlusion events [27]. The insight comes from the fact that depending on whether an object is fully visible throughout the sequence, or is being occluded during some time interval, walls of the object tunnel exhibit different properties. Given that the general problem of occlusion detection, involving multiple, complex, non-rigid moving objects is a very difficult problem to solve, Konrad *et al.* constrain the problem to be the case of one moving object and a static background. In January 2004, Shi *et al.* [55] extended their work by developing a technique based on multiphase level set method. This allowed for an arbitrary number of general motions to be easily incorporated into the framework.

In September 2004, Konrad *et al.* strengthened the notion of explicit occlusion event detection by proposing a framework for the joint object/background segmentation and the detection and modelling of background occlusion and exposed volumes [26]. These novel occlusion and exposure volumes generalise the single-time occlusion field between two images to a continuous event across space and time (Figure 24.) Their motivation to model these events was their potential applications in video compression (occlusion and exposed areas can be thought of as an innovation process, and as such are difficult to predict and expensive to code), and video games (to optimise pixel rendering based on occlusion information.)

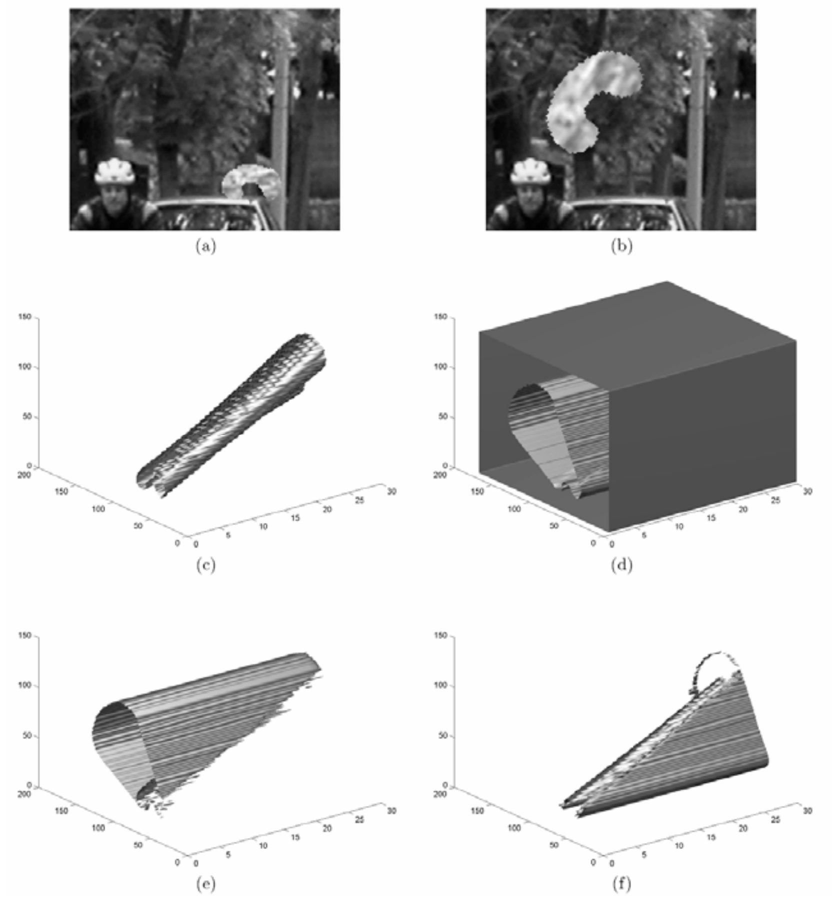


Figure 24 (a) and (b) are frames 1 and 30 from a synthetic test sequence, (c) is the object volume, (d) is the background tunnel, (e) is the occlusion tunnel and (f) is the exposure tunnel.

Also in 2002, the second group at INRS Telecommunications formulated the problem as a Bayesian estimation task, and used the Euler-Lagrange descent equation to minimise a particular energy functional of the segmentation, expressed as a level set partial differential equation [36]. This level set equation was also generalised to the case of multiple motion regions. Some of the results of this work can be seen in Figure 25.

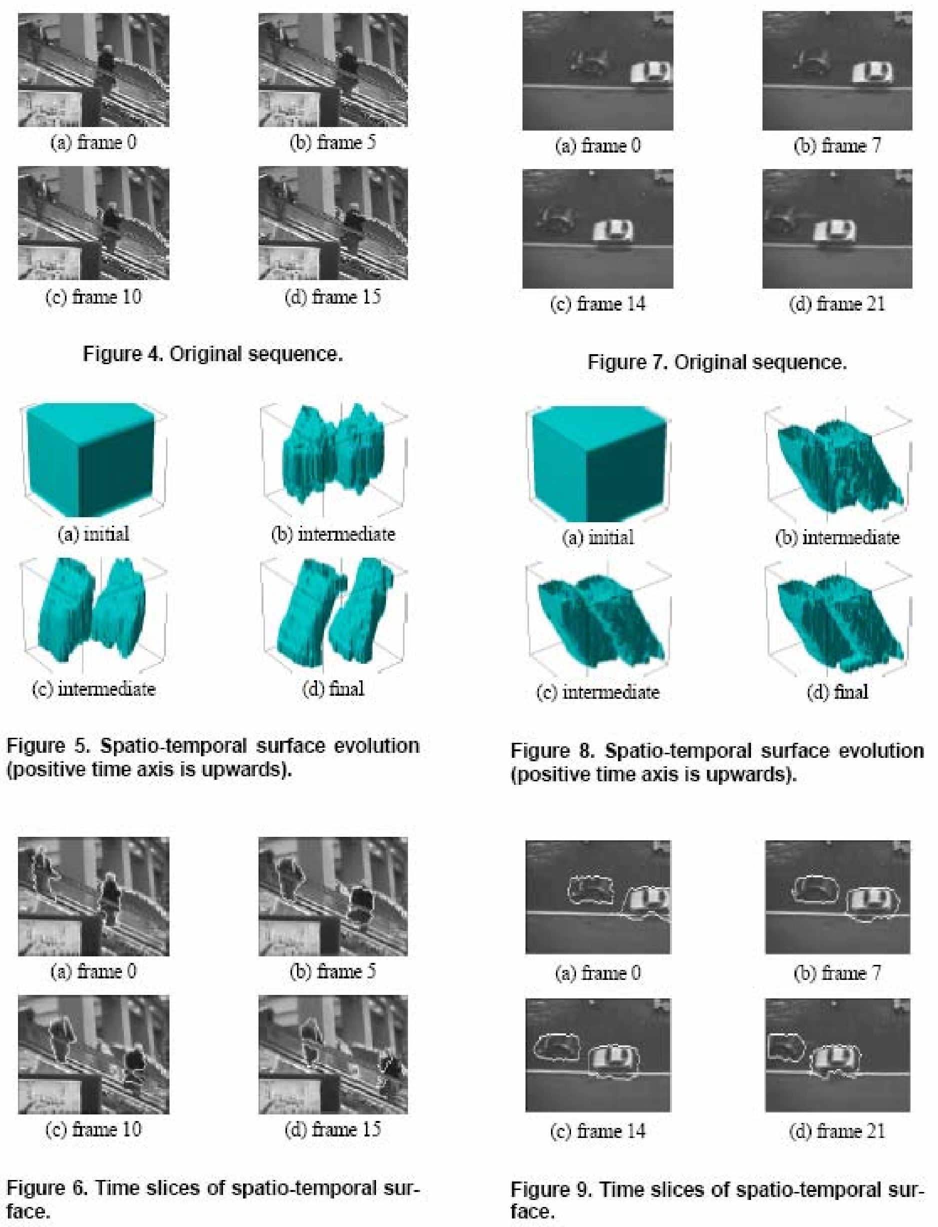


Figure 4. Original sequence.

Figure 7. Original sequence.

Figure 5. Spatio-temporal surface evolution (positive time axis is upwards).

Figure 8. Spatio-temporal surface evolution (positive time axis is upwards).

Figure 6. Time slices of spatio-temporal surface.

Figure 9. Time slices of spatio-temporal surface.

Figure 25 Results of the work of [36] for two image sequences. The top row shows four frames from each sequence, taken with a static camera, the middle row shows the spatio-temporal surfaces being refined iteratively, and the bottom row shows these surfaces projected onto the original frames.

Recently, Feghali *et al.* [14] developed their system to allow for simultaneous camera motion subtraction. Representing the background motion by a parametric model, the estimation of these parameters is not performed as a separate step, but estimated simultaneously using the level set equations. The method performed well for synthetic scenes experiencing modest camera motion.

Conclusion

This report has aimed to outline the most recent techniques used in computer graphics, image processing and computer vision to process image sequences in terms of a 3D spatio-temporal volume. Since the founding work of Aldelson *et al.*, this volume has been analysed by either considering spatio-temporal slices independently, or by considering the spatio-temporal volume in its entirety. Both approaches have their relative advantages and shortcomings. Most notable, slice processing involves exclusively 2D image processing such as curve fitting and 2D Gabor decomposition, which are far cheaper operations than their 3D counterparts. However, by analysing only spatio-temporal slices, the 3D structural information, continuous across all dimensions is lost. If we compare each system with the ultimate goal of image sequence understanding; to spatially segment objects and their motions in the scene at the object description semantic level, each of the approaches fall short of this target. However, what they have achieved are varying degrees of success for a more constrained version of the problem. Nearly all approaches assume constant illumination across a video sequence, since motion is typically detected on the basis of intensity differentials. Also, object motion is commonly constrained either by analysing static scenes with a moving, parameterised camera motion, or by using a static camera photographing a dynamic scene. In both instances, the scenes are usually uncluttered, in which only up to a few objects are considered. In every system reviewed, the objects considered have been assumed to undergo only rigid translations. Furthermore, the majority of techniques do not attempt to solve the joint problem of spatial segmentation and dense motion field estimation.

An intrinsic difficulty in evaluating segmentation algorithms is that the results may currently only be judged through visual examination. These results may well be subjective and inconsistent among different people, and are necessarily qualitative. Unfortunately, the visual performance of each group's systems is presented only as a small series of very similar source texture/synthesised texture pairs. In the past 3D segmentation literature, there appears to be no mention of a standard set of evaluation sequences from which to draw direct comparisons. Also, because the performance of 3D segmentation algorithms typically varies between different types of scenes, allowing the researchers to select their own examples may tempt them to use scenes for which their approach works particularly well.

The major exception to these limitations of the systems, caused by constraining the problem domain, has been through the development of level set techniques. In their latest paper, Feghali *et al.* were able to successfully analyse multiple, rigid moving objects in a cluttered scene with an integrated camera motion subtraction mechanism. A very interesting area of research might be to integrate the modelling of 3D spatio-temporal surfaces (for object segmentation) using the level set methodology with the Shape from Motion mechanism developed by Rodrigues *et al.*, by defining suitable 'interesting points' on the generated surfaces.

Appendix A

Overview of camera motion compensation techniques

It is often necessary to estimate and compensate the motion of a moving camera when analysing an image sequence. When a camera moves, it generates motion over the entire image, and consequently tracking and segmentation problems cannot be solved simply by motion detection. Methods of tracking with a moving camera fall into one of two categories. In the first instance, camera motion parameters are given as input, or that the background scenes have distinctive features or textual properties from which to infer these parameters [Ronsfeld 1998]. Once known, the camera's view coordinates may be translated into world coordinates for affine transformation-invariant segmentation. However, this method clearly restricts the applicability of the segmentation algorithms to arbitrary scenes with unknown camera motion. The second category assumes that background motion is represented by a parametric model, and once an estimate for this is computed, object motion is detected based on motion after compensation for camera motion [Mech 1998, Farin 2001.] Optical flow is commonly used for this purpose, which computes the motion vector of each pixel in an image. For the case of motion through a cluttered 3D scene however, measuring optical flow is problematic because of the high density of depth discontinuities. Rather than measuring velocities at individual points. Mann *et al.* [Mann 2004] recently developed a method which measures a distribution of velocities over local image regions, based on *optical snow*.

Bibliography

[1] Aldelson, E., Bergen, J.R., *"Spatiotemporal energy models for the perception of motion"*, Journal Optical Society of America, vol. 2, 284-299, 1985

[2] Allmen, M., Dyer, C.R., *"Computing relations for dynamic perceptual organization"*, Computer Vision, Graphics and Image Processing: Image Understanding 58, 338-351, 1993

[3] Allmen, M., Dyer, C.R., *"Long-range spatiotemporal motion understanding using spatiotemporal flow curves*", in Proceeding of IEEE International Conference of Computer Vision and Pattern Recognition, 303-309, 1991

[4] Allmen, M., Dyer, C.R., *"Computing spatiotemporal surface flow",* in Proceedings of IEEE International Conference on Computer Vision, 3, 47-50, 1990

[5] Bab-Hadiashar, A., Suter, D., *"Motion Segmentation Using Robust Motion Estimation"*, in Proceeding of Australian Pattern Recognition Society Image Segmentation Workshop, 1996

[6] Baker, H.H, Bolles, R. C., *"Generalizing epipolar plane image analysis on the spatio-temporal surface"*, n Proceedings of the DARPA Image Understanding Workshop, 33-49, 1988

[7] Baker, H.H, Garvey, T.D., *"Motion tracking on the spatiotemporal surface*", in Proceedings of the IEEE Workshop of Visual Motion, 340-345, 1991

[8] Bolles, R.C., Baker, H.H., Marimont, D.H., *"Epipolar-plane image analysis: an approach to determining structure from motion"*, International Journal on Computer Vision, 1, 7-55, 1987

[9] Bouthemy, P., François, E., *"Motion Segmentation and Qualitative Dynamic Scene Analysis from an Image Sequence*", International Journal of Computer Vision, 10(2), 157-182, 1993

[10] Cohen, L., Cohen, I., *"A finite element method applied to the new active contour models and 3-D reconstruction from cross sections*", in Proceedings of International Conference of Computer Vision, 3, 1990

[11] Cohen, L., *"On active contour models and balloons"*, CVGIP, Image Understanding, 211-218, 1991

[12] Cootes, T.F., Cooper, D., Taylor, C.J., Graham, J., *"Active Shape Models - Their Training and Application"*, Computer Vision and Image Understanding. Vol. 61, 38-59, 1995

[13] DeMenthon, D., Megret, R., *"Spatio-temporal segmentation of video by hierarchical mean shift analysis*", Statistical Methods in Video Processing Workshop, 800-810, 2002

[14] Feghali, R., Mitiche, A., "*Tracking with Simultaneous Camera Motion Subtraction by Level Set Spatio-Temporal Surface Evolution*", in Proceedings of IEEE International Conference on Image Processing, vol. 3, 929-932, 2003

[15] Feldmann, I., Eisert, P., Kauff, P., "*Extension of epipolar image analysis to circular camera movements*", in Proceedings of the International Conference on Image Processing, 697-700, 2003

[16] Hamarneh, G., Gustavsson, T., *"Deformable spatio-temporal shape models: Extending ASM to 2D+ Time*", Journal of Image and Vision Computing, vol. 22, pp 461-470, 2001

[17] Hirahara, K., Chenghua, Z., Ikeuchi, K., *"Panoramic-view and epipolar-plane image understangings for street-parking vehicle detection*", in Proceedings of ITS Symposium, 2003

[18] Horn, B. K. P, Schunck, B. G., *"Determining optical flow*", Artificial Intelligence, vol. 17, pp 185-203, 1981

[19] Huang, Y., Palaniappan, K., Zhuang, X., Cavanaugh, J.E., *"Optic Flow Field Segmentation and Motion Estimation Using a Robust Genetic Partitioning Algorithm"*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 17(12), pp 1177-1190, 1995

[20] Jebera, T., Azarbayejani, A., Pentland, A., *"3D Structure from 2D motion*", IEEE Signal Processing Magazine, 16, 1999

[21] Kass, M., Witkin, A., Terzopoulon, D., *"Snakes: Active contour models* ", International Journal of Computer Vision, pp 321-331, 1987

[22] Kawasaki, H., Murao, M., Ikeuchi, K., Sakauchi, M., *"Enhanced navigation systems with real images and real-time information*", International Journal of Computer Vision, vol. 58, Issue 3, pp 237-247, 2004

[23] Kollnig, H., Nagel, H.H., Otte, M., *"Association of Motion Verbs with Vehicle Movements Extracted from Dens Optical Flow Fields*", in Proceeding of third European Conference on Computer Vision, pp 338-347, 1994

[24] Konrad, J., Ristivojevic, M, *"Joint space-time image sequence segmentation based on volume competitions and level sets*", in Proceedings of IEEE International Conference of Image Processing, pp 573-576, 2002

[25] Konrad, J., Ristivojevic, M, "*Joint space-time motion-based video segmentation and occlusion detection using multi-phase level sets*", in Proceedings of SPIE Visual Communications and Image Processing, pp 703-714, 2004

[26] Konrad, J., Ristivojevic, M, "*Joint Space-time image sequence segmentation: Object Tunnels and Occlusion Volumes*", in Proceedings of International Conference on Acoustics, Speech and Signal Processing, pp 9-12, 2004

[27] Konrad, J., Ristivojevic, M., "*Video segmentation and occlusion detection over multiple frames*", in Proceedings of Image and Video Communications and Process, vol. 5022, pp 377-388, 2003

[28] Kuhne, G., Richter, S., Beier, M., "*Motion-based segmentation and contour-based classification of video objects*", in Proceedings of the ninth ACM international conference on Multimedia, 2001

[29] Kühne, G., Richter, S., Beier, M., "*Motion-based segmentation and contour-based classification of video objects*", in Proceedings of the ninth ACM international conference on Multimedia, pp 41-50, 2001

[30] Leymarie, F., Levine, M. D., "*Tracking Deformable Objects in the Plane Using an Active Contour Model*", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 15, Issue 6, pp 617-634, 1993

[31] Li, Y. , Tang, C.-K., Shum, H.-Y., "*Efficient dense depth estimation from dense multiperspective panoramas*", in Proceedings of International Conference on Computer Vision (ICCV), pp 119–126, 2001,

[32] Mandelbaum, R., Salgian, G., "*Correlation-Based Estimation of Ego-Motion and Structure from Motion and Stereo*", in Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999

[33] Mansouri, A., Mitiche, A., Feghali, R., "*Spatio-Temporal Motion Segmentation via Level Set Partial Differential Equations*", IEEE Southwest Symposium on Image Analysis and Interpretation, 2002

[34] Mansouri, A., Mitiche, A., "*Spatial/Joint Space-Time Motion Segmentation of Image Sequences by Level Set Pursuit*", IEEE International Conference on Image Processing, vol. 2, pp 265-268, 2002

[35] Mansouri, A., Mitiche, A., Aron, M., "*PDE-based region tracking without motion computation by joint space-time segmentation*", in Proceedings of International Conference on Image Processing, vol. 2, 2003

[36] Mansouri, A.R., Mitiche, A., "*Spatial/joint space-time motion segmentation of image sequences by level set persuits*", in Proceedings of IEEE International Conference of Image Processing, pp 265-268, 2002

[37] McInerney, T., Terzopoulos, D., "*A finite element model for 3D shape reconstruction and nonrigid motion tracking*", in Proceedings of International Conference of Computer Vision, 4, pp 518-523, 1993

[38] McInerney, T., Terzopoulos, D., "*Topologically adaptable snakes*", in Proceedings of the Fifth International Conference on Computer Vision, pp 840-845, 1995

[39] Mitiche, A., Feghali, R., Mansouri, A., "*Motion Tracking as Spatio-Temporal Motion Boundary Detection*", Journal of Robotics and Autonomous Systems, Vol. 43, pp 39-50, 2003

[40] Murray, D.W., Buxton, B., "*Scene Segmentation from Visual Motion Using Global Optimization*", IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI, 9(2), pp 220-228, 1987

[41] Ngo, C.W., Pong, T.C., Zhang, H.J., "*Motion analysis and segmentation through spatio-temporal slice processing*", IEEE Transactions on Image Processing, pp 341-355, 2003

[42] Ngo, C.W., Pong, T.C., Zhang, H.J., "*Motion-based video representation for scene change detection*", International Journal of Computer Vision, pp 127-142 , 2000

[43] Ngo, C.W., Pong, T.C., Zhang, H.J., Chin, R.T., "*Motion characterization by temporal slice analysis*", IEEE Conference on Computer Vision and Pattern Recognition, 2000

[44] Ngo, C.W., Pong, T.C., Chin, R.T., "*Detection of gradual transitions through temporal slice analysis*", in Proceedings of International Conference of Computer Vision and Pattern Recognition, 1999

[45] Niyogi, S., Adelson, E., "*analyzing and recognizing walking figures in XYT*", in Proceedings of IEEE International Conference of Computer Vision and Pattern Recognition, pp 469-474, 1994

[46] Odobez, J., Bouthemy, P., "*MRF-Based Motion Segmentation Exploiting a 2D Motion Model Robust Estimation*", in Proceedings of IEEE International Conference on Image Processing ICIP'95, pp 628-631, 1995

[47] Ohara, Y., Sagaw, R., Echigo, T., Yagi, Y., "*Gait Volume: Spatio-temporal Analysis of Walking*", The fifth Workshop on Omni directional Vision, Camera Networks and Non-classical cameras, pp 79-90, 2004

[48] Ono, S., Kawasaki, H., Hirahara, K., Kagesawa, M., "*Ego-motion estimation for efficient city modeling using epipolar plane range image analysis*", in ITSWC2003, 2004

[49] Otsuka, K., Horikoshi, T., Suzuji, S., "*Image velocity estimation from trajectory surface in spatiotemporal space*", In Proceedings of International Conference on Computer Vision and Pattern Recognition, pp 200-205, 1997

[50] Peng, S.L., Medioni, G., "*Interpretation of image sequences by spatio-temporal analysis*", Workshop on Visual Motion, pp 344-351, 1989

[51] Peng, S.-L., "*Temporal slice analysis of image sequen*ces", in Proceedings of Computer Vision and Pattern Recognition, pp 283-288, 1991

[52] Phillip, A., Laplante, P.A., "*Real-time Imaging - Theory, Techniques and Applications*", IEEE Press, New York, 1996

[53] Rav-Acha, A., Peleg, P., "*A Unified Approach for Motion Analysis and View Synthesis* ", 2nd International Symposium on 3D Data Processing, Visualization, and Transmission, pp 717-724, 2004

[54] Rodrigues, R., Fernandes, A., Overveld, K., Ernst, F., "*Reconstructing depth from spatiotemporal curves*", in Proceedings on the 15th International Conference on Vision Interface, 2002

[55] Shi, Y., Konrad, J., Karl, W.C., "*Multiple motion and occlusion segmentation with a multiphase level set method*", Symposium on Electronic Imaging, Visual Communications and Image Processing, pp 18-22, 2004

[56] Shi, Y., Konrad, J., Karl, W. C., "*Multiple motion and occlusion segmentation with a multiphase level set method*", in Proceedings of International Conference on Visual Communications and Image Processing, 2004

[57] Swaminathan, R., Kang, S.B., Criminisi, A., Szeliski, R., "*On the Motion and Appearance of Specularities in Image Sequences*", in Proceedings of European Conference in Computer Vision, 2002

[58] Wang, J., Adelson, E., "*Representing Moving Images with Layers*", IEEE Transactions on Image Processing, 3(5) pp 625-638, 1994

[59] Willis, C., "*Video Stack Image Analysis*", Edinburgh University MSc Dissertation, 2004