Available online at http://www.journalcra.com

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

International Journal of Current Research
Vol. 3, Issue, 4, pp.050-056, April, 2011

ISSN: 0975-833X

REVIEW ARTICLE

PIXEL WISE SEGMENTATION OF MOTION COHERENT PARTICLES

***Pandi Selvi, P. and Mahesh, K.**

Department of Computer Science and Engineering Alagappa University, Karaikudi, Tamailnadu, India.

ARTICLE INFO

Article History:
Received 10th January, 2011
Received in revised form
14th February, 2011
Accepted 11th March, 2011
Published online 17th April 2011

Key words:
Ensemble clustering,
Motion segmentation,
Object-based video segmentation,
Point tracking,
Video coding,
Reality-based 3-D models.

ABSTRACT

The survey describes an approach for object –oriented video segmentation based on motion coherence. Using a tracking process ,2-D motion patterns are identified with an ensemble clustering approach. Particles are clustered to obtain a pixel-wise segmentation in space and time domains. The limitation of the segmentation method concerning with complex 3-D spatial motion can be solved by using reality-based 3-D models. The reality-based 3-D models are produced with range-based, image-based or CAD modeling techniques. Each 3-D model can contain different levels of geometry and need therefore to be semantically segmented and organized in different ways. Index Terms-Ensemble clustering, motion segmentation, object-based video segmentation, point tracking, video coding, reality-based 3-D models.

INTRODUCTION

Motion segmentation is an important preprocessing step in many computer vision and video processing tasks, such as surveillance, object tracking, video coding, information retrieval, and video analysis. The definition of object in a video segmentation framework is related to the concept of region homogeneity, and different applications require different region homogeneity criteria. In video coding, for example, segmentation is frequently used to explore the data redundancy in time. In this context, an object region that retains its characteristics (e.g., color or texture) along the sequence can be considered homogeneous and redundant. Thus, even if the object region moves along the temporal sequence, the region representation remains the same, i.e., redundant, within the object motion boundaries. In 3-D motion segmentation, the concept of object is related to actual objects existing in 3-D space that do not change their 3-D characteristics over time. The concept of object in spatio-temporal segmentation is different, since an object is represented by a spatio-temporal tunnel formed by a sequence of 2-D projections, each 2-D projection obtained at a time of an object in 3-D space. On the other hand, when camera translations and depth variations are small compared to the distance of the camera to the

Corresponding author: selvikrish.selvi@gmail.com.

scene objects, simpler 2-D motion models can become attractive . In 2-D parametric motion segmentation, a small number of parameters is needed to describe the object motion, making the motion segmentation more robust to noise. 2-D motion segmentation also has received attention, since it still has some open issues and it is suitable for some important video processing tasks like video coding, where a simple representation is important, and in general the semantic aspects of the scene are less relevant .

In the context of motion estimation, the literature can be divided in two classes of methods: direct methods and feature-based methods . Direct methods recover the unknown parameters directly from measurable image quantities at each pixel in the image, solving two problems simultaneously: 1) the motion of the camera and/or objects of the scene, and 2) the correspondence of every pixel.Feature-based methods minimize an error measure that is based on distances between a few corresponding features, while direct methods minimize a global error measure that is based on direct image information collected from all pixels in the image. For this reason, direct methods are sometimes called dense methods in the literature. Thus, in this work we refer to any motion estimation / segmentation method that yields correspondence / classification for each pixel as "dense", even if the motion estimation/ segmentation core is guided only by a sparse set of features. It is important to observe that with direct methods the pixel correspondence / classification is performed directly with the measurable image quantities at each pixel, while in feature-based methods this is done indirectly, based on independent feature measurements in a set of sparse pixels.

An important property of the direct methods is that they can sucessfully estimate global motion even in the presence of multiple motions and/or outliers. On the other hand, feature-based methods initially ignore areas of low information, resulting in a problem with fewer parameters to be estimated, with good convergence even for long sequences. Therefore, Mitiche *et al.* proposed to segment moving objects by detecting the tunnel delimited by motion discontinuities in the spatio-temporal domain. Feghali and Mitiche further extended this idea to also handle moving cameras, and later Sekkati and Mitiche proposed a 3-D direct motion segmentation method with a similar approach . They formulated the problem as a Bayesian motion partitioning problem, and approached the corresponding Euler-Lagrange equations as a level-set problem. Cremers and Soatto proposed a multiphase level-set method to segment a video using spatio-temporal surfaces (tunnels) that separate regions with piecewise constant motion.

A limitation of segmentation methods based on motion discontinuities (motion boundaries) is that they tend to fail in frames where these boundaries are not evident, or do not exist. Ristivojevic and Konrad also proposed a spatio-temporal segmentation method based on the level-set approach, where they defined the concepts of occlusion volumes (i.e., background regions that become occluded) and exposed volumes (background regions that become visible). As occlusion and exposed areas are difficult to predict, new efficient compression techniques could be developed by estimating occlusion and exposed areas *a priori.* A characteristic shared by most variational methods is that they rely on motion models defined *a priori.* If the data do not fit these models well, the methods tend to fail—except when special conditions can be assumed, like static background, number of objects known *a priori,* etc.

Feature-based methods for motion segmentation usually consist of two independent stages: 1) feature selection and/or correspondence and 2) motion parameter estimation . The second stage often is performed through factorization methods. Also, homogeneous regions of a frame may present none or few features, making the motion estimation/segmentation difficult (or even impossible) in large areas of the video frames. Eventually, we can obtain consistent object segmentation by combining several partial informations about an object. The idea of merging object segmentation information from several parts of a sequence was proposed by Geldon as a probabilistic multiple hypothesis tracking (PMHT) approach. The authors propose to track an object over the whole image sequence, by combining

partial object segmentations previously computed in different parts of the sequence. This is done by modeling the motion and geometry of the objects, and these models are combined assuming smooth trajectories, and are used to eliminate ambiguities caused by occlusions and incorrect detections. However, the object motion and/or geometry modeling accuracy depends on how well the models fit the data, besides the trajectory constraints preclude the application of this approach in videos with discontinuous object trajectories, which is common in sequences obtained with hand-held cameras, for example. Here we present a new approach for video object segmentation where objects are defined as nonoverlapping regions (at pixel level) in the spatio-temporal domain. Our approach combines the advantages of dense and feature-based methods, as described next. Initially, correspondences in time of sparse points (i.e., particles) are computed, so that long-range motion patterns can be identified. However, instead of computing point correspondences independently (as done in many feature-based methods), neighboring particles are treated as they were linked, reducing the chance of occurring outliers and avoiding the aperture problem .Moreover, the density of sampled points (i.e., particles) is adaptive, and denser particle distributions are used in regions where precision is more important (for example, in motion boundaries), saving computation without neglecting homogeneous regions.

To compute particle correspondences in a video sequence, we use the approach proposed by Sand and Teller ,which relies on particles that are located with sub-pixel precision. After the particle correspondences are computed, particles are clustered in each frame of the sequence. The individual particle clusterings at frame level, are then further grouped in larger sets of particles associated to different frames, according to an ensemble clustering strategy. Finally, a dense video frame representation (i.e., a pixel-wise representation) of the final clustering is obtained.And in the case of 3-D complex modeling in which it has the drawback of over segmentation it can be rectified with the help of Reality-Based 3-D Modeling. Reality-based surveying techniques (e.g. photogrammetry, laser scanning, etc.) employ hardware and software to metrically survey the reality as it is, documenting in 3D the actual visible situation of a site by means of images ,range-data ,Reality-Based 3D Modeling and Segmentation of the aforementioned techniques .Non-real approaches are instead based on computer graphics software or procedural modeling and they allow the generation of 3D data without any metric survey as input or knowledge of the site.

Range-Based3D Reconstruction

Optical range sensors like pulsed (TOF), phase-shift or triangulation-based laser scanners and stripe projection systems have received in the last years a great attention, also from non-experts, for 3D documentation and modeling purposes. During the surveying, the instrument should be placed in different locations or the object needs to be moved in a way that the instrument can see it under different viewpoints. Successively, the 3D raw data needs errors and outliers removal, noise reduction and the registration into a unique reference system to produce a single point cloud of the surveyed scene or object. The registration is generally done in two steps: (i) manual or automatic raw alignment using targets or the data itself and (ii) final global alignment based on Iterative Closest Points (ICP) or Least Squares method procedures.

Image-Based3D Reconstruction

Image data require a mathematical formulation to transform the two-dimensional image measurements into three- dimensional information. Image-based modeling techniques (mainly photogrammetry and computer vision) are generally preferred in cases of lost objects, monuments or architectures with regular geometric shapes, small objects with free-form shape, low-budget project, mapping applications, deformation analyses, etc.The dense 3D reconstruction step can instead be performed in a fully automated mode with satisfactory results

CAD-Based 3D Reconstruction

This is the traditional approach and remains the most common method in particular for architectural structures, constituted by simple

geometries. In addition, each volume can be either considered as part of adjacent ones or considered separated from the others by non-visible contact surfaces. Using CAD packages, the information can be arranged in separate regions, each containing different type of particles, which help the successive segmentation phase. The proposed method is general in the sense that it does not rely on motion models, does not impose trajectory constraints and segments multiple objects of arbitrary shapes, without knowing the number of objects *a priori*. Instead of motion boundaries, the segmentation is guided by the consistent motion behavior of sample points of the frames. The proposed method potentially has the ability to generate occluded and exposed volumes, since motion patterns are discovered and associated to each moving region, and voxels of the spatio-temporal volumes are classified as belonging to object volumes, occluded volumes or exposed volumes. The proposed approach generates a simple scene representation, adequate for object video coding, and also delivers a more redundant and temporally persistent partition of the scene than direct video segmentation methods and motion prediction strategies.

METHOD OVERVIEW

The structure of the proposed coherent motion segmentation approach can be divided in three main parts.

Estimation of Particle Trajectories
Segmentation of Particle Trajectories
Dense Segmentation Extraction

Estimation of Particle Trajectories

It concerns the selection and tracking of a set of points of the scene (namely, particles). This stage takes as input the original video frames, and returns as output a set of particles and their respective trajectories. During the estimation of particle trajectories, the particles whose correspondent point locations in the scene suffer occlusion are eliminated, and new particles are created in regions that become newly visible along the video sequence.

Segmentation of Particle Trajectories

It deals with the segmentation of particle trajectories, so that particles moving coherently are grouped together. This stage takes as input the particles trajectories computed in the first stage, and returns labels for all the particles as outputs, representing the motion segmentation of frame regions according to the particle trajectories. The segmentation of particle trajectories can be divided in four steps.

Clustering of 2-frame motion vectors:

In this step, clusterings of particles are performed with displacement motion vectors taken from pairs of frames. Only neighboring frames are considered (1, 2, and 3 unit time distances), and clusterings are computed in an independent way. For each pair of frames, the input to this step is the position of particles in each frame, and the output is a set of particle clusterings and their labels, valid for each pair of frames considered.

Ensemble clustering of particles

Here, all the clusterings computed in the previous step are processed simultaneously to produce a unique division of the full set of particles in sub-sets of particles in coherent motion, called meta-clusters; several sets of clustering labels are taken as input to this step, and a unique set of segmentation labels (several particles in coherent motion share the same segmentation label) are returned as output.

Meta-clustering validation

In this step, particles that were segmented in the previous step are compared to meta-cluster prototypes in terms of motion and spatial position to detect incorrectly labeled particles and, when this occurs, particles are re-labeled. A set of segmentation labels is taken as input, and a corrected set of segmentation labels is returned as output.

Spatial filtering

In this step, outliers are eliminated and groups of adjacent particles that are not significant. The particle labels are analyzed spatially, and links between particles are created to define spatial adjacency in each frame. Small groups of adjacent particle labels that are not significant are then re-

assigned. This step takes as input a set of particle labels, and returns as output a filtered set of particle labels.

Dense Segmentation Extraction

The proposed motion segmentation method is the dense segmentation extraction. This stage takes as input the original video frames, the segmentation labels returned by the second stage, as well as the particle positions returned by the first stage, and returns as the output the corresponding segmentation labels for each pixel of each frame of the video sequence. This is equivalent to the segmentation of a spatio-temporal volume in several tunnels. The dense segmentation extraction is done by creating implicit functions for each particle, based on motion and spatial position. This representation of motion segmentation through tunnels can be employed to obtain efficient motion predictions for video coding applications. All the stages of the proposed approach are processed sequentially. Every stage is performed for the entire video before going to the next stage. Thus, the proposed motion segmentation method can not be used in online applications, without video partitioning.

3D Model Segmentation

The segmentation of a polygonal model consists in the decomposition of the 3D geometry into sub-elements which have generally uniform properties. But in most of the applications the user intervention is still mandatory to achieve more accurate results. The main reasons that limit the automatic reconstruction of semantic models are related to:

- the definition of a target model which restricts object configurations to sensible building structures and their components;
- the geometric and radiometric complexity of the input data and reconstructed 3D models;
- data errors and inaccuracies, uncertainty or ambiguities in the automatic interpretation and segmentation;
- the reduction of the search space during the interpretation process.

The 3D geometry and the related semantic information have to be structured coherently in order to provide a convenient basis for simulations, urban data analyses and mining, facility management, thematic inquiries, archaeological analyses, policies planning, etc.

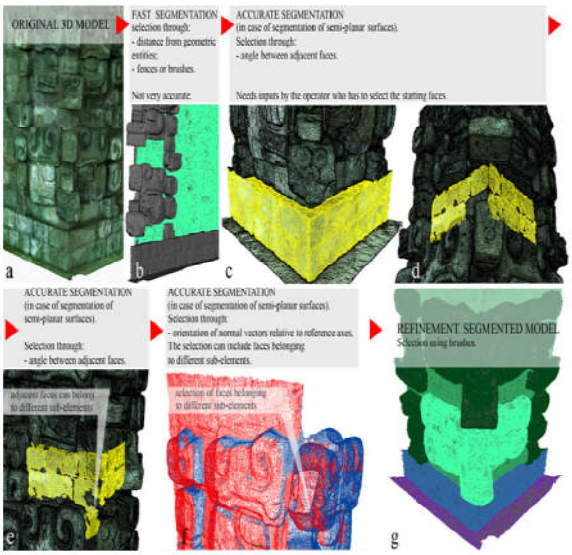

Fig. 2. Example of automated segmentation of complex and detailed polygonal model

a fast but inaccurate segmentation can be improved with geometric constraints and manual refinements to separate the narrative elements In the literature, the most effective automated segmentation algorithms are based on 3D volumetric approaches, primitive fitting or geometric segmentation methods. While the former two approaches segment meshes by identifying polygons that correspond to relevant feature of the 3D shape, the latter segments the mesh according to the local geometrical properties of 3D surface. The methodology specified here also follow these concepts and uses a combination of automated and interactive segmentation tools according to the 3D model and its complexity (Fig.2). Furthermore the segmentation is generally performed according to rules or specifications differs for each project.
The segmentation procedure performs:

- an automatic geometric separation of the different mesh portions using surface geometric information and texture attributes;
- a manual intervention to adjust the boundaries of the segmented elements
- an assisted annotation of the sub-elements that constitute the segmented 3D model.

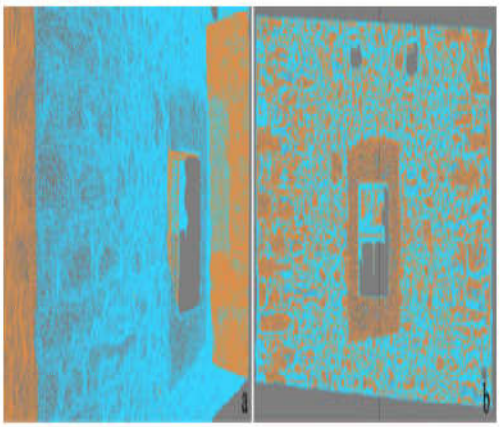

Fig. 3. Segmentation of a stone wall of a castle (Fig.2-f) with small surface irregularities. a) result derived aggregating the main surface orientations; b) result of the segmentation following planar adjacent faces.

Faces can be separated and grouped using constraints such as inclination of adjacent faces, lighting or shading values. The surface normals are generally a good indicator to separate different sub-elements.This can be quite useful in flat areas with very low geometric discontinuities where the texture information allow to extract, classify and segment figures or relevant features for further uses (Fig. 5).

For complex geometric models constituted by detailed and dense meshes, the manual intervention is generally required. Another aspect that has to be considered during the geometric segmentation of complex and fully 3D models is related to the possibility of subdividing only visible surfaces or to build complete volumes of sub-element models, modeling also non- visible closure or transition surfaces.

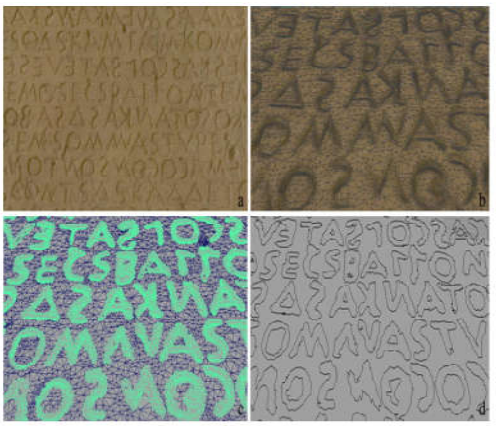

Fig. 4. a) Image of the law code in Gortyna (Fig.2-e) with symbols of ca 3-4 mm depth; b)close view of the 3D textured polygonal model; c) automatic identification of the letters using geometric constraints; d) final segmentation and vectorialization of the letters

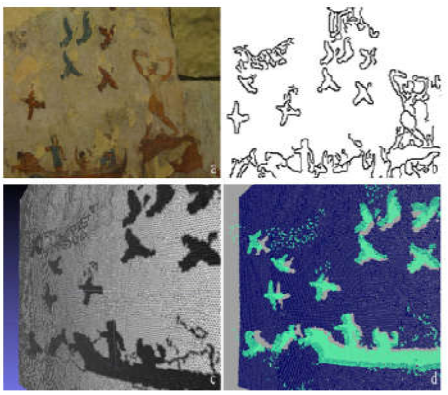

Fig.5. a) The 3D model of an underground frescoed Etruscan tomb, b) filters to detect edges on the texture information of the 3D model; c,d) ease of lighting transitions and final segmentation of the polygonal model

CONCLUSION

A study has been made in this paper for the identification of coherent motion in adaptively sampled videos. This technique provides a new way of linking low-level information in videos to high-level concepts that can be employed directly in video coding. The proposed particle segmentation method uses ensemble clustering to combine particle clusters obtained for adjacent frames, allowing the identification of long-range

motion patterns, which we represent as spatio-temporal volumes called tunnels. The identification of long-range motion patterns is crucial to take full advantage of temporal redundancy in segmentation- based video coding. Another limitation of the proposed segmentation method concerns the type of motions that are better handled with this approach. Since we generate clusters based on similarity of the 2-D projected motion vectors, sequences with pronounced perspective effects and/or with complex 3-D spatial motion tend to be over-segmented. In order to overcome this drawback Reality-Based 3-D modeling has been proposed here which enable us to semantically segment complex reality-based 3-D models.

REFERENCES

Gruen, A, 2008. Reality-based generation of virtual environments for digital earth. *International Journal of Digital Earth* 1(1)

Vosselman, G., Maas, H.G, 2010. Airborne and terrestrial laser scanning, p. 320. CRC Press, Boca Raton.

Yin, X., Wonka, P., Razdan, A, 2009. Generating 3d building models from architectural drawings: A survey. *IEEE Computer Graphics and Applications*, 29(1), 20–30.

Guidi, G., Remondino, F., Russo, M., Menna, F., Rizzi, A. and Ercoli, S, 2009. A multi-resolution methodology for the 3d modeling of large and complex archaeological areas. *International Journal of Architectural Computing,* 7(1), 40–55.

Remondino, F., El-Hakim, S., Girardi, S., Rizzi, A., Benedetti, S. and Gonzo, L, 2009. 3D virtual reconstruction and visualization of complex architectures - The 3d-arch project. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 38(5/W10), on CD-ROM.

Agarwal, S., Snavely, N., Simon, I., Seitz, S. and Szelinski, R, 2009. Building Rome in a day. In: Proc. ICCV 2009, Kyoto, Japan.

Hiep, V.H., Keriven, R., Labatut, P., Pons, J.P, 2009. Towards high-resolution large-scale multiview stereo. In: Proc. CVPR 2009, Kyoto, Japan.

Barazzetti, L., Remondino, F. and Scaioni, M, 2010. Automation in 3D reconstruction: results on different kinds of close-range blocks. In: ISPRS Commission V Symposium Int. Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Newcastle upon Tyne, UK, 38(5).

Chen, X., Golovinskiy, A, and Funkhouser, T, 2009. A Benchmark for 3D Mesh Segmentation. *Proc. ACM Transactions on Graphics,* 28(3), 12.

****