3D Representation of Videoconference Image Sequences Using VRML 2.0

Ioannis Kompatsiaris and Michael G. Strintzis

Information Processing Laboratory
Electrical and Computer Engineering Department
Aristotle University of Thessaloniki
Thessaloniki 54006, Greece
ikom@dion.ee.auth.gr

Abstract. In this paper a procedure for visualisation of videoconference image sequences using Virtual Reality Modeling Language (VRML) 2.0 is described. First image sequence analysis is performed in order to estimate the shape and motion parameters of the person talking in front of the camera. For this purpose, we propose the K-Means with connectivity constraint algorithm as a general segmentation algorithm combining information of various types such as colour and motion. The algorithm is applied "hierarchically" in the image sequence and it is first used to separate the background from the foreground object and then to further segment the foreground object into the head and shoulders regions. Based on the above information, personal 3D shape parameters are estimated. The rigid 3D motion is estimated next for each sub-object. Finally a VRML file is created containing all the above estimated information.

Keywords : *virtualised reality; model-based image sequence analysis; Virtual Reality Modeling Language.*

1 Introduction

Modeling of 3D scenes from 2D images has been the focus of considerable attention in literature [1–3]. Video production, realistic computer graphics, multimedia interfaces and medical visualisation are some of the applications that may benefit by exploiting the potential of 3D model-based schemes. This usually requires the estimation of the 3D object shape, 3D object motion and object surface.

While much work has been done in the signal processing community, in order to extract those parameters from real sequences, there is a lack in interportability and commonly accepted formats. For the time being, MPEG-4 standard working towards this direction is still in the development process [4]. On the other hand, VRML file format for describing 3D worlds has been developed by the computer graphics society, to represent mainly synthetic 3D environments [5]. VRML is a widely accepted format and it is used on the World Wide Web for dynamic interaction with 3D worlds.

The incorporation of real scenes into such environments is called *Virtualised Reality* [6]. Like Virtual Reality (VR), Virtualised Reality also immerses the viewer in a virtual environment. The two differ, however, in the way the virtual worlds model are constructed. VR environments are typically created using simple CAD models and lack fine detail, especially in the texturing part. Virtualised Reality, by contrast, automatically constructs the virtual model from images of the real world, preserving the visible detail of the real-world images. Furthermore, other aspects of the real world can be estimated and be added to the virtualised environment, such as motion. In [7] 2D motion extracted from real images was used to animate synthetic 2D objects.

In this paper a procedure for visualisation of a videoconference image sequence in VRML 2.0 is described. First image sequence analysis is performed in order to extract the parameters of the real world. Then these parameters are converted into VRML format and a moving 3D representation of the image sequence can be viewed by any VRML compliant browser. The visualisation offers enhanced telepresence to the viewer, since a 3D representation of the scene is created. Furthermore, the user inside the VRML browser can interact with the scene (e.g. change the lighting conditions) and synthetic objects can be integrated in the virtualised environment. A segmentation scheme based on the K-Means clustering algorithm is used, in order to extract the 3D shape and motion parameters.

The paper is organised as follows. In the following Section the K-Means with connectivity constraint is described, while in Section 3 a spatiotemporal filter used to derive motion information is given. In Section 4 the 3D shape parameters are estimated. In Section 5 the 3D model is used for rigid 3D motion estimation. The visualisation process using VRML is described in Section 6. Finally in Section 8 conclusions are drawn.

2 The K-Means with Connectivity Constraint Algorithm

Clustering based on the K-Means algorithm is a region segmentation method which is widely used [8]. The main problem of K-Means clustering is that the regions produced are not solid and there are parts of a region contained in others. In order to solve this problem, we propose an extended K-Means algorithm : the K-Means-with- connectivity-constraint algorithm. During the clustering procedure, before a pixel is assigned to a specific object, it is checked whether the objects remain solid and the merging criterion is also affected by this factor. Furthermore we combine information from different sources in the K-Means algorithm for efficient segmentation.

More specifically, each pixel is subdivided according to its position on the image (x, y) and a function describing characteristic pixel properties such as colour and motion :

$$F(x,y) = a_c\tilde{I}(x,y) + a_m\tilde{V}(x,y) \ ,$$

where $\tilde{I}$ is a normalised version of the initial image between 0 and 1 and $\tilde{V}$ is also a normalised version of the picture produced by taking the frame differences of