Chapter 10
Synthetic Video Generation for Evaluation of Sprite Generation

Yi Chen
University of Alabama in Huntsville, USA

Ramazan S. Aygün
University of Alabama in Huntsville, USA

ABSTRACT

Sprite generation is the process of aligning, warping, and blending of pixels that belong to an object in a video. The evaluation of the correctness of a sprite is usually accomplished by a combination of objective and subjective evaluations. Availability of ground-truth image would help mere objective evaluation. In this paper, the authors present video generation from an image based on various camera motion parameters to be used as ground-truth for the sprite evaluation. This paper introduces a framework for evaluation of sprite generation algorithms. Experiments under the proposed framework were performed on the synthetic videos of different camera motion patterns to reveal the components of the sprite generation algorithm to be improved.

INTRODUCTION

The term "sprite" refers to the composition of pixels that belong to a video object in a video. Sprite generation (Lu, 2003) is the process of generating sprites for objects in videos. The most common object for sprite generation is the background, and the corresponding sprite is usually referred as the background sprite. It was used for video compression under MPEG-4 video standard (Sikora, 1997). Sprite coding is only supported by MPEG-4 Main Profile and requires the availability of the sprites. In MPEG-4, a sprite may be generated for each object, and objects are layered on top of each other. Instead of sending the complete background scene for every frame, sprite coding encodes and transmits static background sprite once. The individual frames can be regenerated from the sprite with the support of motion compensation. Thus, sprite coding leads

DOI: 10.4018/978-1-4666-1791-9.ch010

to good subjective quality with very low bitrates (Jinzenji, 2001).

Sprite generation has been also studied as background extraction (Lai, 2008), photo stitching (Baudisch, 2007; Brown, 2007; Baudisch, 2005; Zomet, 2006), and panoramic image generation (Farin, 2008). Sprite generation research has gained significant attraction since late 1990s due to its application in many domains such as object-based coding (Wantanabe, 2001), video compression (Jinzenji, 2001), video indexing (Grammalidis, 1999), virtual environments (Jaillon, 1994), object tracking (Lin, 2002), and security surveillance (Cheng, 2007). Sprite generation is mostly applied to the background in the video. In the rest of the paper, we will use the term sprite for the background sprite unless stated otherwise. Since background scene may not be captured in a single frame of a video, sprite is generated by correctly overlapping between sequential frames. Sprite is usually generated when there is a significant global motion and the background regions eventually become visible. Therefore, sprite generation algorithms process every frame (usually in order) to incorporate occluded backgrounds as they become visible and newly visible backgrounds. Sprite generation is mostly applied in order sequentially for all frames in the video, since the global motion is limited for consecutive frames. The consecutive frames are aligned and the newer one is blended into the sprite.

The three main steps of sprite generation are global motion estimation, warping and blending. Global motion estimation aims at calculating the motion between two (consecutive) frames. For background sprite, it is usually assumed that global motion estimation corresponds to camera motion if the moving objects are not too large. Global motion estimation is the most computation intensive component of sprite generation. Warping maps coordinates of pixels in one frame to coordinates of pixels in the other frame without changing pixel values. In other words, it aligns the frames that are processed. Blending determines what the pixel values should be after aligning the frames (e.g., averaging).

Global motion estimation depends on the motion models used for the estimation. The most common motion models are 2D motion models, such as perspective, affine, or translational models, due to their performance. Since cameras capture 3D environments where object distances to camera vary and some local objects may move in a different direction from the camera motion, the 2D models may not be 100% accurate to model the environment. Hence, 2D global motion estimation algorithms can only approximate the actual global motion. If the estimated global motion is close enough to the actual motion, the sprite can still be generated by using enhanced methods such as long-term global motion estimation (Smolic, 1999). If the global motion is not estimated correctly, neither warping nor blending may produce satisfactory results.

The evaluation of sprite generation is composed of subjective and objective evaluation steps. An expert looks at the sprite and the video and determines whether it is an appropriate sprite or not in the subjective evaluation. In the objective evaluation, the typical measure is peak signal-to-noise ratio (PSNR) (Li, 2003). The PSNR (Smolic, 2003; Sheikh, 2006) is used to calculate the difference between the generated frame and original frame. The visual quality was maintained despite decrease in PSNR (Smolic, 2003). Sheikh et al. (Sheikh, 2006) state that the correlation of mean squared error (Li, 2003) and PSNR is not a tight indication of human judgment quality. The problem for sprite evaluation is that the PSNR value can be ∞ even for an incorrectly generated sprite if the frames are concatenated without any overlapping. In the objective evaluation, the original frames can be regenerated exactly the same although the sprite is not correct at all. However, the generated sprite is not correct. (To check the correctness of the sprite, we regenerated the original frames, and the original frames can be regenerated exactly the same although the sprite is not correct at all).

Related Content