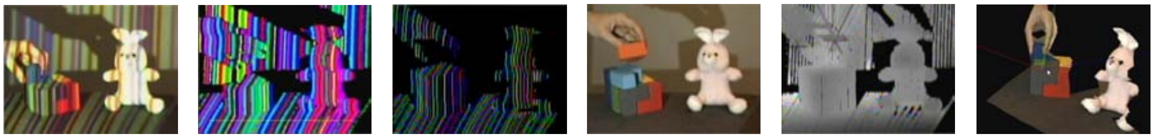
Real-Time 3D Video

Luiz Velho
IMPA

Marcelo Bernardes Vieira
IMPA

Asla Sá
IMPA

Paulo Cezar Carvalho
IMPA



(a) slides (b) patterns (c) boundaries (d) texture (e) depth (f) 3D

Figure 1: System pipeline.

The advent of digital video caused a technological revolution that changed in many ways the audio-visual communications. The digital format is, by its very nature, robust to degradations and well suited to computer processing. As a consequence, digital video had a great impact on the Television and Film industries. Now, it is shaping the new media, with the Internet and wireless channels.

The development of the initial generations of digital video focused on technological foundations, system issues and standards. In this respect, the result was a significant improvement of *evolutionary* aspects, such as image resolution. But the data type remained the same – i.e., color and sound information.

The next generation of digital video has to bring innovation related to *revolutionary* aspects. This will be the result of incorporating new types of data into the media. Depth information is certainly the most natural candidate of data type to augment digital video. Not only it is consonant with the human perceptual system, but it also facilitates scene analysis by computers to extract information at higher levels.

We will call *3D video* a time-varying image with color + depth information. The first challenge to 3D video is to develop an acquisition device. The most obvious choice would be a system based on a pair of cameras and passive stereo methods. However, fully general stereo is an ill-posed inverse problem which is very hard to solve – and real-time requirements make matters even worse. Another choice, would be a system based on a camera / projector and active stereo methods. This option has the advantage of simpler and robust constrained stereo algorithms, but the price is that a pattern of light has to be projected on the scene.

In this poster, we describe a real-time 3D video system that is based on active stereo. Our main motivation to develop this system is to create a plataform for investigating the issues that will be posed by the next generation of digital video and how it will shape up new media.

Our system is a complete plataform for 3D video, consisting of an aquisition device, data processing, transmission and visualization modules. The system generates 3D video in real-time (30 fps) from a wide range of scenes.

The acquisition device employs active stereo and is composed of callibrated, synchronized video camera and projector. The data processing module extracts depth information from structured light code. The transmission module performs I/O as well as data stream compression and decompresion. The visualization module renders 3D video using dynamic point-based geometry. Here, we emphasize the 3D video capture aspects of the system.

Scene depth reconstruction is obtained by triangulation of camera–projector correspondences. We have extended the (b,s) BCSL method [Sa et al. 2002], to work with dynamic scenes. The (b,s) BCSL is a structured light color boundary code which consists of a sequence of s patterns of stripes with b colors. We adopted a $(2,6)$ code – i.e., two patterns with six colors (R,G,B,C,M,Y). This configuration gives the best compromise between temporal-space coherence and code lenght.

The system actually employs a sequence of complementary pattern pairs – i.e., $S_1\overline{S_1}S_2\overline{S_2}$, where S_k is a stripe pattern and $\overline{S_k}$ its color complement. This scheme makes the code identification very robust and boundary localization more precise. We also note that complementary color pattern pairs correspond to a constant white color (in fact, we verified experimentally that if the patterns are projected at 120Hz they are percieved as white light). The complementary (b,s) BCSL code makes possible to capture, at the same time, geometry and texture of the 3D scene. It also makes the construction of shadow masks a trivial task.

Figure 1 gives an overview of the 3D processing pipeline. The sequence $S_1\overline{S_1}S_2\overline{S_2}$ of stripe patterns is projected (Fig. 1a); From each pair $S_k\overline{S_k}$, stripe colors and boundaries are inferred (Figs. 1b and 1c); Then, boundary codes are derived from matching patterns S_{ij} Texture and shadow mask is obtained from combining images $\overline{S_k\overline{S_k}}$ (Figs. 1d); Depth is computed by triangulation of boundary correspondences at camera and projector, followed by interpolation (Fig. 1e); Finally the 3D video is assembled, transmitted and displayed (Fig 1f).

The key observation that explains why the (b,s) BCSL coding works well for dynamic scenes is that the projected stripe boundaries change very little, even when objects in the scene move fast. Boundary displacement is only due to object motion towards the camera.

Callibration is a very important component of the system. It includes both geometric and fotometric callibration of the camera and projector. The use of probabilistic methods based on apriori knowledge makes the system more robust and reliable. We also exploit spatial and temporal coherence.

The acquisition device is built with off-the-shelf NTSC video equipment. This has many advantages, such as good cost-performance, ease of synchonization, compatibilitiy and many options of distribution channels. We are now beginning to develop applications that use our 3D video system.

References

SA, A., CARVALHO, P. C., AND VELHO, L. 2002. (b, s)-BCSL: Structured light color boundary coding for 3d photography. In *Proceedings the 7th VMV*.