Contact information:

Claudio Pinhanez, Frank Nielsen, Kim Binsted

Sony Computer Science Laboratory
Takanawa Muse Building
3-14-13 Higashi-Gotanda, Shinagawa-ku
Tokyo 141-0022 Japan

pinhanez@media.mit.edu, nielsen@csl.sony.co.jp, kimb@csl.sony.co.jp

Projecting Computer Graphics on Moving Surfaces: A Simple Calibration and Tracking Method

We are currently investigating the possibilities of transforming every surface in a space into a display and, in particular, of projecting graphics on movable surfaces. We present here a simple method to convert tracking points from a camera into positions on the image to be projected. The main advantage of our method is the simplicity of the calibration process that does not require the knowledge of the intrinsic parameters (focal length, dimensions of the imaging device) neither of the camera nor of the projector.

Camera and Projector Calibration

The relationship between points observed on a planar surface from two different cameras is known to be a *homography* [1]. A homography is a 3x3 matrix defining a linear application in the projective space that, for a given planar surface of the real world, maps all projected points in one camera's image into the other camera's image.

The fundamental observation is that from a geometrical point of view, "ideal" pinhole projectors and cameras are identical (see fig. 1). Let H denote the homography that relates the image of the projector image frame to the camera image frame. This means that a 2d-point homogeneous coordinates on the camera image $\bar{c} = \left(x_c/z_c\,, y_c/z_c\right)$ matches a 2d point $\bar{p} = \left(x_p/z_p\,, y_p/z_p\right)$ on the projector image as follows:

$$p = \begin{pmatrix} x_p \\ y_p \\ z_p \end{pmatrix} = Hc = H \begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix}$$

A homography is completely defined if the projection of four 3d points of the world on both image planes is known. To determine the homography between a camera and a projector, we need simply to obtain the four needed points while manually aligning a projection of the surface with the real surface (see fig. 1).

The homogeneous coordinates of the four points to be projected, $p_i = \left(x_p^i, y_p^i, 1\right)$, $i = 1,2,3,4$, are determined arbitrarily, although making sure that the points are visible and there is a way to move the real surface so it aligns with the projection. Then, we consider the homogeneous coordinates of the four points on the camera image as sensed by the tracking system, $c_i = \left(x_c^i, y_c^i, 1\right)$, $i = 1,2,3,4$. Taking the matrices corresponding to these two sets of four points, $P = \left(p_1^T\ p_2^T\ p_3^T\ p_4^T\right)$ and $C = \left(c_1^T\ c_2^T\ c_3^T\ c_4^T\right)$, we obtain $P = HC$, whose solution is

$$H = PC^T\left(CC^T\right)^{-1}$$

During run-time, we simply take a point in the camera image $c = (x_c, y_c, 1)$, project through the homography H obtaining $p = Hc = (x_p, y_p, z_c)$, and compute the position on the projector's image plane, $(x_p/z_p, y_p/z_p)$.

Surprisingly, this calibration step is numerically stable even with only four points and can be done, in practice, in a few seconds. We believe that the stability is also related to the fact that in our experiments the projection centers of the camera and the projector are close to be aligned. Notice that there is no need to determine neither the camera's intrinsic parameters nor the projector's.

Tracking the Projection Surface

In our experiments we have used plain markers on the projection surface. In particular, we employed infrared LEDs that can be easily tracked by a camera with an infrared filter. However, if we move the mask too quickly, we observe that the projected image "falls behind" the moving surface. That is, there is a "shifting" effect where the observations at discrete time t on the camera image $c(t)$ are displayed by the projector at time $t + dt$ using the estimate at time t, $p(t) = Hc(t)$. To reduce the "shifting" problem we employ a predictive Kalman filter [2], that estimates the most likely position of every point at time $t + dt$, using the equations of dynamics as the underlying model of the Kalman filter, as shown in fig. 1. The parameter dt, corresponding to the average delay between sensing and displaying, is determined experimentally. The Kalman filtering approach proved to be very effective in our experiments.

References

[1] O. Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint*. The MIT Press, Cambridge, Massachusetts. 1993.

[2] A. Gelb (ed.). *Applied Optimal Estimation*. The MIT Press, Cambridge, Massachusetts. 1974.

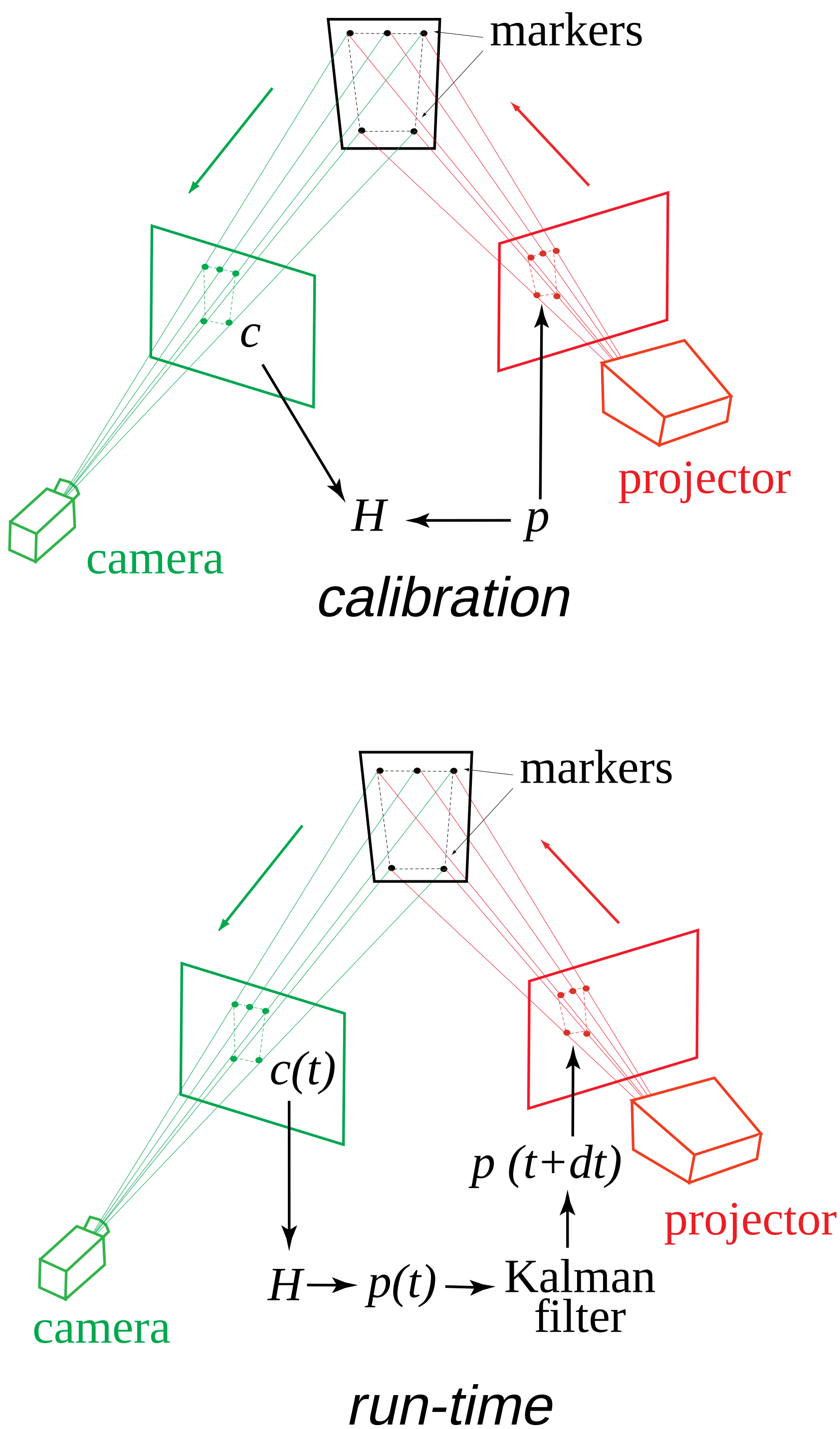

Figure 1. Calibration process and run-time system.