Tracking People from Above

Algorithm

Frames are read from the video stream with interval 0.3 seconds.

The bounding box in the first frame is positioned by hand. Size of the box is a little larger than head size.

SURF features are detected in the frame within the bounding box.

Point Tracker (KLT algorithm) is initialized with locations of these features.

Make one step of point tracking using the next frame. Some points may be lost in this operation.

Estimate the geometric transformation between the old points and the new points and eliminate outliers. More points may be lost here.

Apply this transformation to the bounding box. This is the position of bounding box for the next iteration.

Speeded-Up Robust Features (SURF)

Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool, "SURF: Speeded Up Robust Features", Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346--359, 2008

SURF is scale- and rotation-invariant interest point detector and descriptor. It outperforms other methods with respect to repeatability, distinctiveness, and robustness, yet can be computed and compared much faster.

To detect interest points, SURF uses the determinant of Hessian matrix - a 2x2 matrix of convolutions of the Gaussian second order derivatives with the given image. It can be computed with 3 integer operations using a precomputed integral image.

Its feature descriptor is based on the sum of the Haar wavelet response around the point of interest. These can also be computed with the aid of the integral image.

SURF descriptors have been used to locate and recognize objects, people or faces, to reconstruct 3D scenes, to track objects.

KLT Point Tracker

Lucas, Bruce D. and Takeo Kanade. "An Iterative Image Registration Technique with an Application to Stereo Vision,"Proceedings of the 7th International Joint Conference on Artificial Intelligence, April, 1981, pp. 674–679.

Tomasi, Carlo and Takeo Kanade. Detection and Tracking of Point Features, Computer Science Department, Carnegie Mellon University, April, 1991.

Shi, Jianbo and Carlo Tomasi. "Good Features to Track," IEEE Conference on Computer Vision and Pattern Recognition, 1994, pp. 593–600.

The method is based on the equation that relates the image gradient **g**, the inter-frame displacement **d**, and the difference h between image intensities:

$\mathbf{g} \cdot \mathbf{d} = h$

The image gradient **g** can be estimated from one image, while the difference h is computed from both.

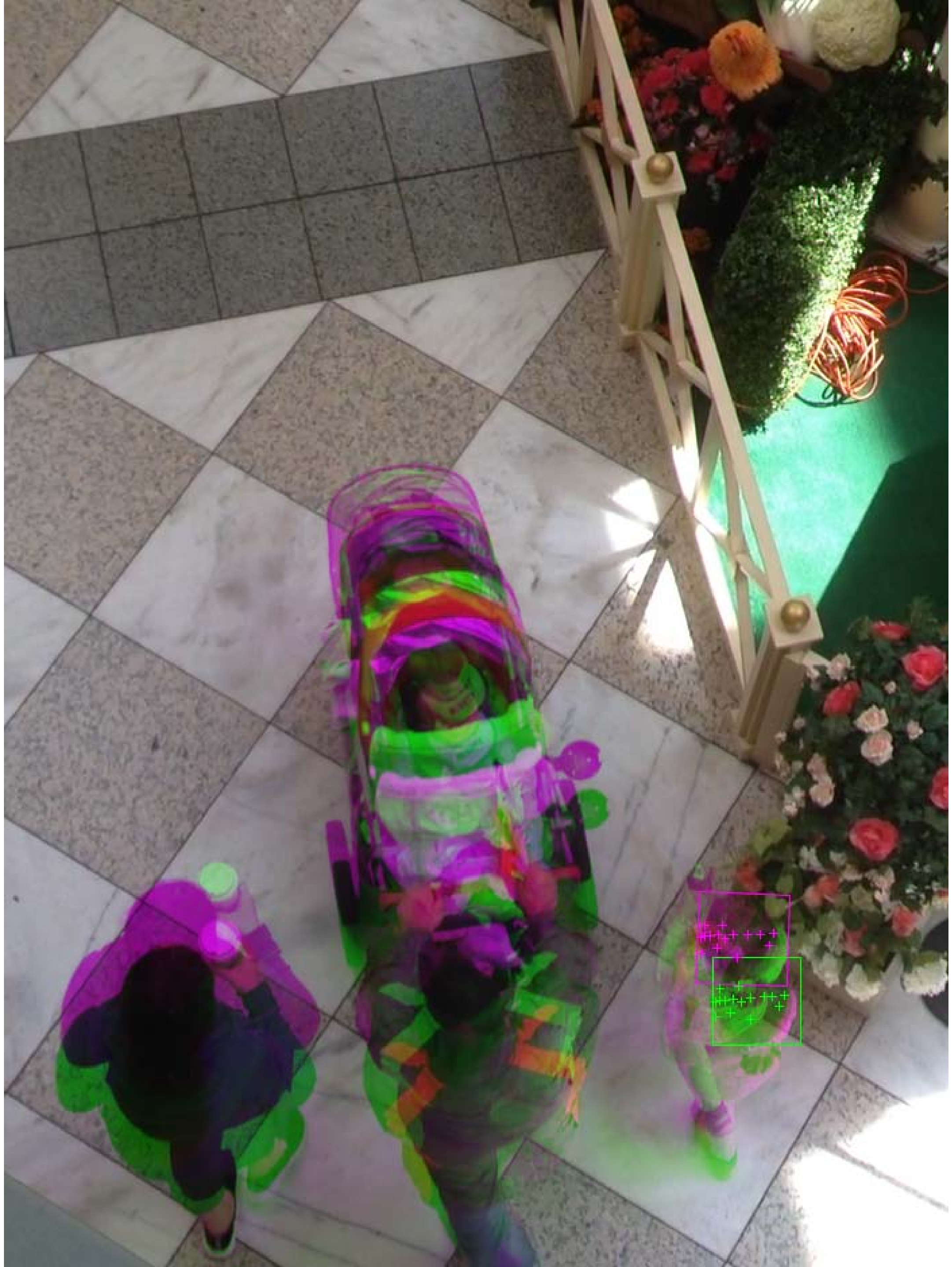
This is a scalar equation with a two-dimensional unknown **d**. The best value for **d** can be chosen as the one that minimizes the error in a window around each tracked point.

Frame 100.0 sec.

Frame 100.3 sec.

Frame 100.6 sec.

Frame 100.9 sec.



Frame 101.2 sec.

Frame 101.5 sec.

Frame 101.8 sec.

Frame 102.1 sec.

Frame 102.4 sec.

Frame 102.7 sec.

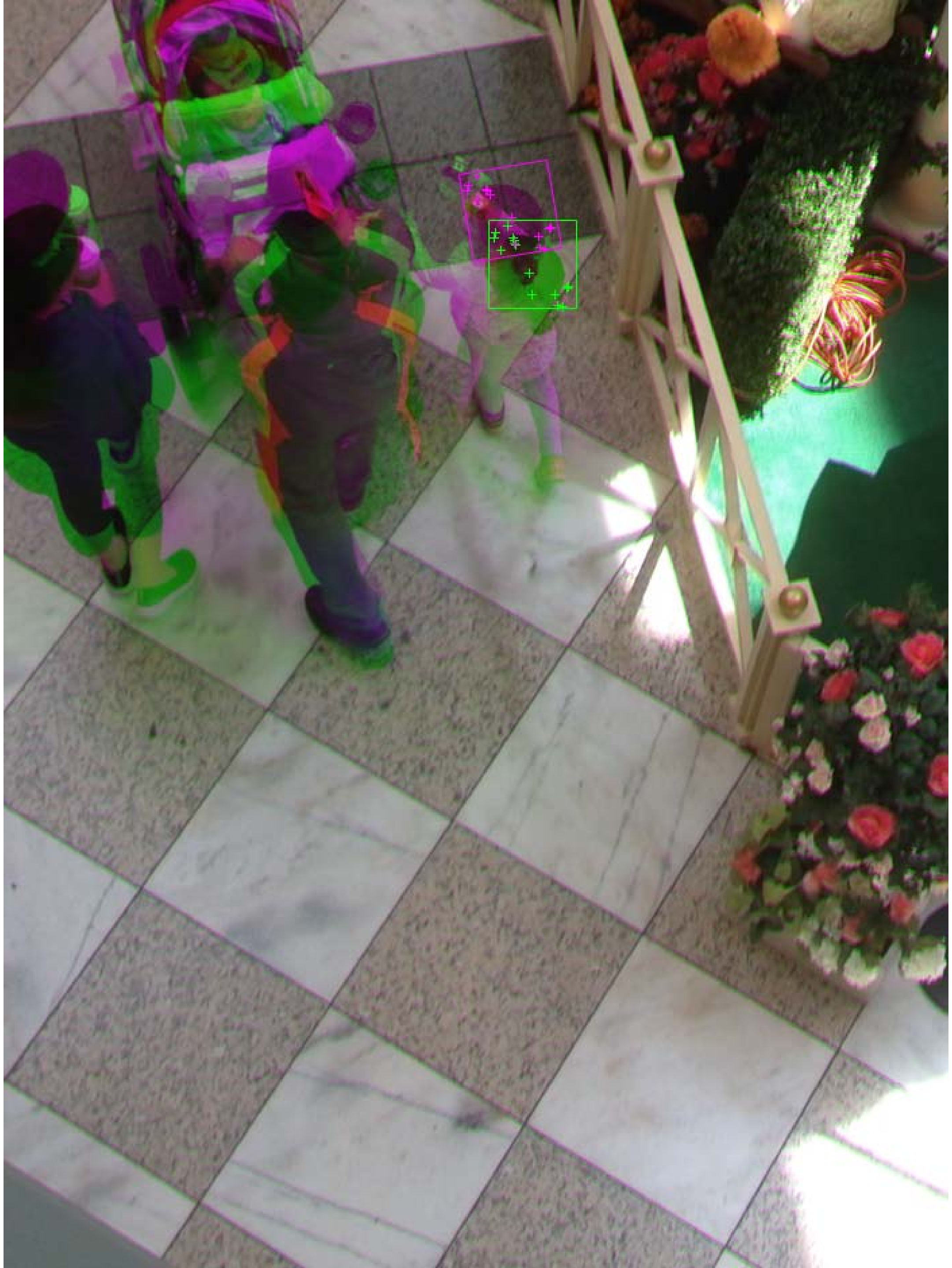
Frame 103.0 sec.

Frame 103.3 sec.

Frame 103.6 sec.

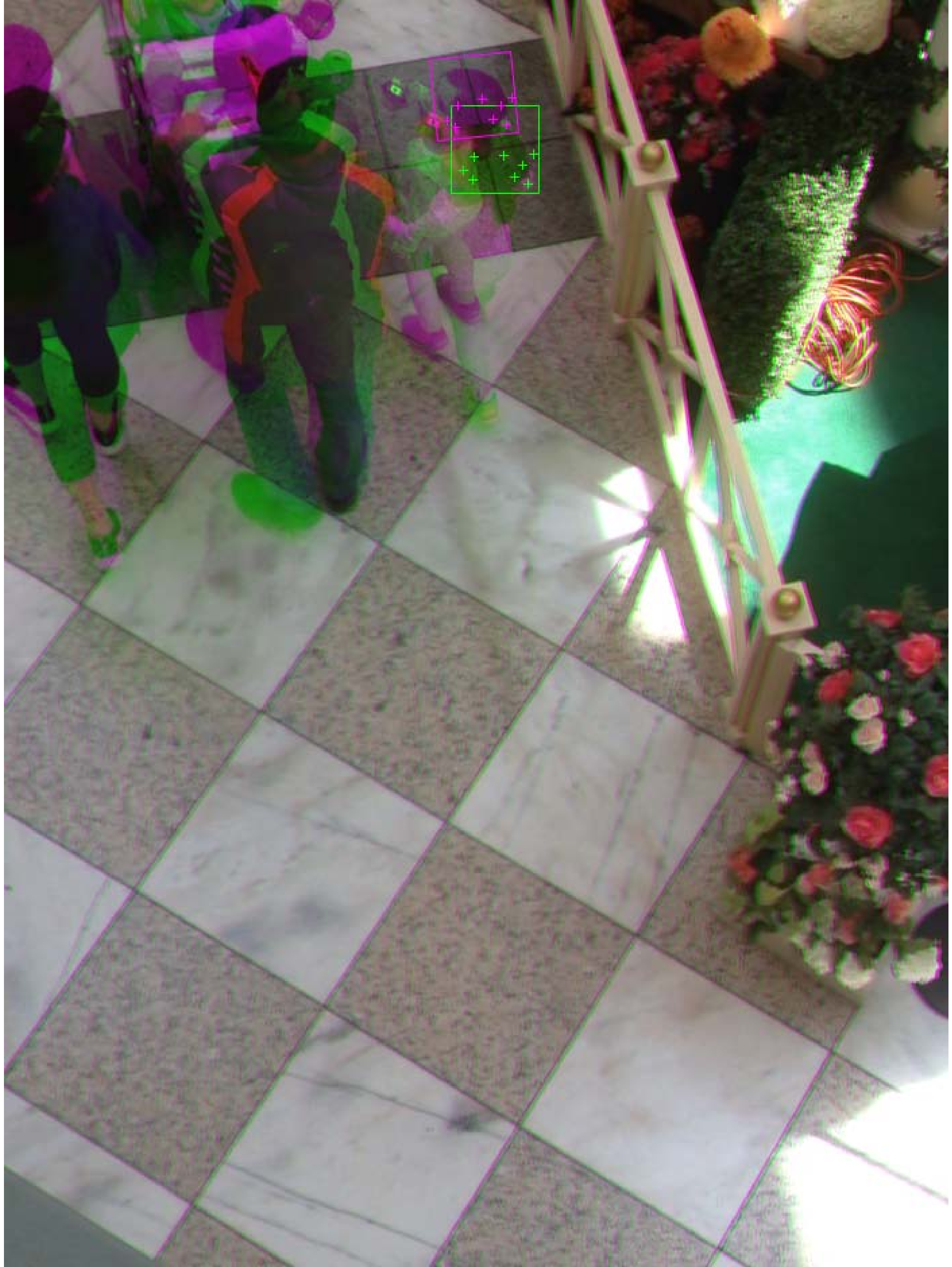
Frame 103.9 sec.

Frame 104.2 sec.

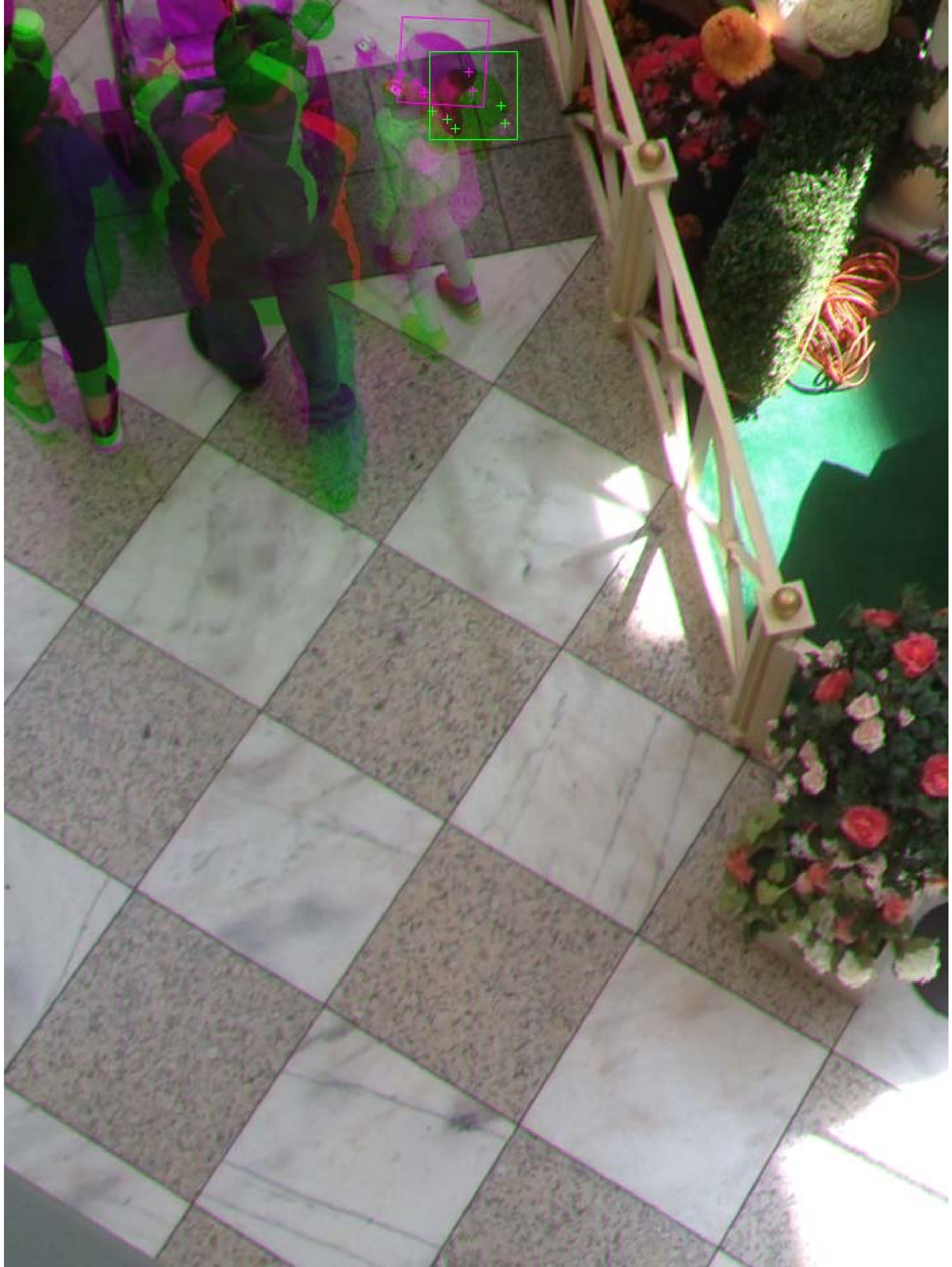


Frame 104.5 sec.

Frame 104.8 sec.



Frame 105.1 sec.



Frame 105.4 sec.

Frame 105.7 sec.

Collage

The algorithm can track a head from the bottom of the frame to the top of the frame.

If the camera fields of view overlap, this means that it can “hand over” the head to the next camera.

Collage

The algorithm can track a head from the bottom of the frame to the top of the frame.

If the camera fields of view overlap, this means that it can "hand over" the head to the next camera.

Collage

The algorithm can track a head from the bottom of the frame to the top of the frame.

If the camera fields of view overlap, this means that it can “hand over” the head to the next camera.

Collage

All heads tracked in parallel

Collage

The algorithm can track a head from the bottom of the frame to the top of the frame.

If the camera fields of view overlap, this means that it can "hand over" the head to the next camera.

Collage

Doesn't have to be a head.
Other objects can be tracked too.

Starting points

In all previous examples the starting point for the algorithm was determined manually.
The next presentation “Finding Faces and Tracking Objects” describes how to do it automatically.