

# Stereo-matching Techniques

Automation in photogrammetry is largely made possible by stereo-matching techniques that enable (semi-)automatic aero-triangulation and creation of digital elevation models. Matching: I know what it is, I know what it does, but how does it work?

The aim of matching is to identify corresponding phenomena in two or more sets. In photogrammetry the sets are the left and right image of a stereo pair, and the problem in corresponding them is to trace and locate in the right image the conjugate of a point in the left image. Ever since images could be stored in a computer as pixels stereo-matching algorithms have been being developed there are many and new ones emerge on a regular basis. The methods may be categorised into two broad classes depending on how the image is approached. From the signal-processing perspective an image is regarded as a set of grey values or colour representing the intensity of reflected electromagnetic signals. But an image may also be seen as a representation of features present in object space where each feature, such as the corner of a building, a road crossing or tree, is represented by an irregularly shaped group of pixels.

In the signal-based approaches, also called area or intensity-based, correspondence is sought using intensity values in a regularly shaped patch. A target patch the size, for example, of 9x9 pixels, is defined and shifted over a search patch in the other image the size of which depends on how well the approximate location of the conjugate point is known. For each position a similarity measure is computed, for example, the normalised cross-correlation, resulting in a connected set of similarity measurements, one for each pixel. The highest similarity value determines the corresponding patches, their centre pixels selected as corresponding points. Acceptance of the match depends on whether the similarity value exceeds a predefined threshold; the value can be selected in a heuristic way or, when using normalised cross-correlation, on a statistically sound basis using student t-testing. Sub-pixel accuracy can be achieved by fitting a function, for example a second-order polynomial, through the correlation values and then determining the maximum of that function.

Correlation techniques allow at best a linear difference (gain and shift) between the intensity values of the left and right image, but a shift only in geometry. Since geometric differences do exist as a result of differences in exterior orientation and presence of relief, these are tackled through an iterative least-squares approach. These usually model the geometric differences as affine transformations. However, gain comes at a cost: the approximate location of the conjugate point has to be known accurately in advance, even down to the level of a few pixels. This problem can be coped with along two lines. The first by establishing an approximate match with feature-based matching, whereby first points, line or areas are detected in both images using differential operators such as Marr-Hildreth, Sobel, Moravec or Förstner. Next, attributes are assigned to the features, such as average and variability of grey values. Knowing the search range, corresponding features in the left and right images are found by comparing attributes. A consistency check is then performed, based on the assumption of smooth object surfaces, to remove faulty assignments. The location of features serves as an approximation for least-squares matching.

Another way to tackle the approximate value problem is by adopting a multi-resolution approach in which an image pyramid is created with at its base the original, full-size images and at subsequent higher levels images generated from uniting 2x2 pixels. This may be repeated until an image of just one pixel remains. By selecting a hierarchical level that best reflects the approximate position of the conjugate point least-squares matching is carried out at that level. The resulting correspondence is now used to track the matching down through the image pyramid until the original image has been reached.

High computation load requires reduction of search space. This is achieved by using information on relative position and orientation of the stereo pair and characteristics of the terrain topography. The first enables use of epipolar geometry, which reduces matching to a 1D problem; the latter might avoid starting at too high a level in the image pyramid.