RGB-D object recognition and localization with clutter and occlusions

Federico Tombari, Samuele Salti, Luigi Di Stefano
Computer Vision Lab
DEIS, University of Bologna
Bologna, Italy

This work demonstrates object recognition robust to clutter and occlusions in 3D data represented by range maps enriched with color information, also allowing for recognition of multiple instances of the model to be found. Our approach compares each scene to a library of models, each model being represented by a set of different views of the model itself. For each model and scene, the stages sketched in the diagram of Figure 1 are applied, which will be now briefly described.

Given a model and a scene, first features are extracted and described on each model view (offline) and on the scene (at run-time) by means of, respectively, SURF detector applied on the texture image and Color-SHOT (CSHOT) [1], a recently proposed 3D descriptor, which represents an extension to RGB-D data of the SHOT descriptor [2]. SHOT and C-SHOT will be now briefly outlined.

The SHOT descriptor [2] relies on the definition of a repeatable local Reference Frame (RF) based on the Eigenvalue Decomposition of the scatter matrix of the neighborhood of a point. Given the local RF, an isotropic spherical grid is defined to encode spatially well localized information, i.e. to define a signature structure. For each sector of the grid, the angles between the normal of the central point and that of each point falling in that sector are accumulated in a histogram. The final descriptor results from the juxtaposition of these histograms. Motivated by the increasing availability of 3D sensors capable of delivering both shape and color information, in [1] an extension to the SHOT descriptor is proposed. This novel descriptor, dubbed CSHOT and aimed at feature matching in 3D data enriched with color, adds to geometrical information (angles between normals of the features, as in SHOT) also color information (color differences between points computed in the CIELab space).

Following the description stage, each model view is efficiently matched against the scene by means of *kd-trees*, determining scene-to-model correspondences. By applying the Hough-based 3D Object Recognition scheme proposed in [3] a subset of geometrically-coherent correspondences is selected on each view. As for this approach, each model feature point is associated with its relative position with respect to the centroid of the model, so that each corresponding scene feature can cast a vote in a 3D Hough space to accumulate evidence for possible centroid position(s) in the current scene. This enables simultaneous voting of all feature correspondences within a single 3-dimensional Hough space. To correctly cast votes according to the actual pose(s) of the object(s) being sought, we rely on the local RFs associated with each pair of corresponding features.

The view with the highest number of selected correspondences is chosen as the best one: if this number is higher than a threshold, the object is detected on the scene, and its pose determined by applying a final RANSAC and Absolute Orientation stage [4] on the remaining correspondences, so as to yield a Rotation matrix and Translation vector that aligns the best model view to the scene.

With our current multi-thread implementation, processing RGB-D data acquired with a Microsoft *Kinect* sensor, around 8 seconds are required to go through all stages of the algorithm on an Intel *Core2 Quad* processor, with a library including 3 models, 4 views per model. A demo video can be found online [1].

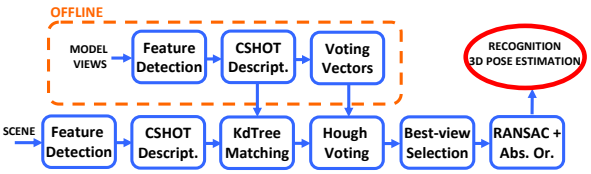

Figure 1: Stages of the proposed algorithm.

References

[1] F. Tombari, S. Salti, and L. Di Stefano, "A combined intensity-shape descriptor for texture-enhanced 3d feature matching," in *ICIP, submitted*, 2011.

[2] F. Tombari, S. Salti, and L. Di Stefano, "Unique signatures of histograms for local surface description," in *Proc. ECCV*, 2010.

[3] F. Tombari and L. Di Stefano, "Object recognition in 3d scenes with occlusions and clutter by hough voting," in *Proc. PSIVT*, 2010.

[4] B. Horn, "Closed-form solution of absolute orientation using unit quaternions," *J. Optical Soc. of America A*, vol. 4, no. 4, 1987.

[1] `www.vision.deis.unibo.it/demos/rgbdDemo.avi`