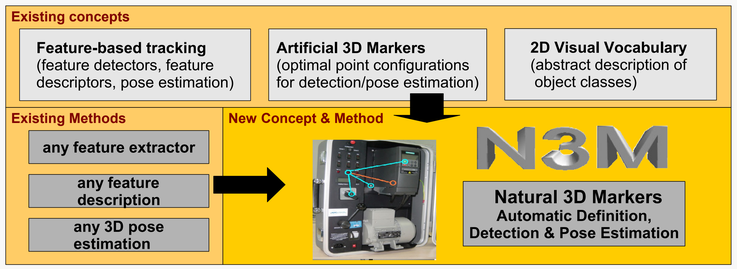
N3M: Natural 3D Markers for Real-Time Object Detection and Pose Estimation

Stefan Hinterstoisser, Selim Benhimane, Nassir Navab
Chair for Computer Aided Medical Procedures (CAMP) - TUM, Germany

Overview

Idea

Automatic recovery of local minimal subsets of features defining a Natural 3D Marker (in analogy to existing "3D optical markers"), resulting in abstract description of objects ("3D sentences") in terms of such N3M entities ("3D vocabulary") suitable for automatic detection and pose estimation



Results

- Fast detection requiring only few inliers,
- Efficient outlier removal,
- Pose estimation based on N3Ms, i.e. quasi optimal point configurations

Learning N3Ms: Definition/Selection

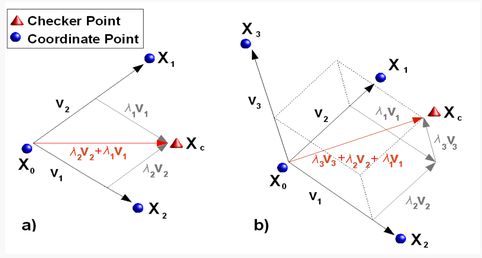
Definition

An N3M is a minimal set of 3D *points* defining a local coordinate system and one 3D *checker point* expressed in this local coordinate system allowing for geometric consistency check.

Pre-processing

- Learn most stable Feature Points over varying viewpoints
- Recover their 3D structure
- Select feature points covering large image regions to increase robustness against partial occlusion
- Compute visibility set for each feature point, i.e. set of views from which the feature point can be detected, to avoid grouping feature points that can not be easily detected under similar viewpoints

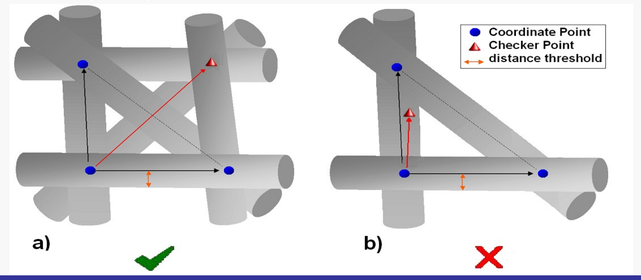
Learn Natural 3D Markers



I. Generate all possible N3Ms in local Neighborhoods:
- Form sets of 4 coplanar (Fig.a) and 5 non-coplanar points (Fig.b)
- Select one of the points in each set as checker point
- Define a local coordinate system using all other points in the set
- Represent the checker point in this local coordinate system

Learning N3Ms: Filtering

II. Remove ill-conditioned N3Ms for pose estimation and self-verification: configurations where the points are too close to each other or they are almost collinear e.g. in Fig. b. have to be removed



Learning N3Ms: Pt. classifiers & Checker Pts.

III. Point Classifiers: For each Feature Point learn a point classifier e.g. based on the Randomized Trees [V. Lepetit et al., CVPR 2005]

IV. Local Coordinates of Checker Points: Compute and store the Coordinates $(\lambda_1,\lambda_2,\lambda_3)$ of the 3D Checker Point Xc in the local Coordinate System $\{X_0,V_1,V_2,V_3\}$:

$$X_c = X_0 + \sum_{i=1..3} \lambda_i V_i$$

where X_i, $i \in \{0,1,2,3,c\}$ are the 3D coordinates of the Feature Points, $V_i = X_i - X_0$, $i \in \{1,2,3\}$ and X_0 is the origin of the local coordinate system.

Run Time Stage

I. Matching: Match all 3D Feature Points X_i with the 2D image points x_j using the trained classifier.

II. Self-verification: Each N3M can be *self-verified* independently on the other N3Ms. A score function allows to measure how well the 3D points of one N3M are matched with their 2D projections. This is defined based on the position of the Checker Point in N3M's local coordinate system:

$$x_c = PX_c \approx AX_c = x_0 + \sum_{i=1..3} \lambda_i v_i$$

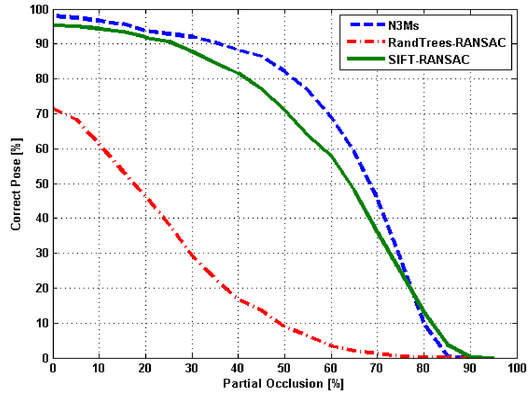
$$f = \left\| (x_c - x_0) - \left(\sum_{i=1..3} \lambda_i v_i \right) \right\|$$

where P represents the unknown camera projection matrix, A is its approximation as a fronto-parallel projection, x_i $i \in \{0,1,2,3,c\}$ are the 2D feature points in the current image, corresponding to the 3D points of an N3M and $v_i = x_i - x_0$, $i \in \{1,2,3\}$.

III. Voting and Similarity Measurement: The few remaining outliers are removed by a) a voting scheme (N3Ms are voting for each other – if enough N3Ms vote for one N3M then this N3M is considered as properly matched) or b) by computing a similarity measure (only for planar N3Ms) between the area of the current image enclosed by the 2D feature points and the texture of the 3D model enclosed by the corresponding N3M.

IV. Compute final pose: the final pose of the object is computed based on the correspondences of all remaining N3Ms.

Performance



The additional step in the Detection/Pose Estimation Pipeline (here with voting scheme) improves the overall performance significantly, resulting in:
- high invariance to partial occlusion
- high invariance to cluttered backgrounds
- fast outlier elimination
- selection of well-conditioned point configurations for pose estimation

Real Time Object Detection and Pose Estimation