Histogram of Oriented Cameras - A New Descriptor for Visual SLAM in Dynamic Environments

Katrin Pirker
kpirker@icg.tugraz.at
Matthias Rüther
ruether@icg.tugraz.at
Horst Bischof
bischof@icg.tugraz.at

Institute for Computer Graphics and Vision
Graz University of Technology
Graz, Austria

1 Motivation

Simultaneous localization and mapping (SLAM) is a fundamental prerequisite in autonomous mobile robotics. Most exisiting solutions assume a static environment (e.g. [1], [3] or [2]) where the environment map is constructed once through adding all incoming sensor data over time. This results in ever growing maps, data association problems and may lead to false environment representations especially in dynamic scenes.

We address the following problems:

- How to handle short- and long-term environmental changes.
- How to balance map size.
- How to improve data association in a dynamic environment.

2 Feature Descriptor

To overcome these problems we propose a new three-dimensional descriptor called Histogram of Oriented Cameras (HOC) to add spatial visibility information to local map features. The HOC keeps track how often a feature has been observed from a specific location. The descriptor consists of k concentric spheres, where each sphere is approximated by an icosahedron of m faces as shown in Figure 1(a). Each bin holds an integer n, which is increased in case of a positive observation from a sensor resting in this pyramid frustum, and decreased if the sensor does not observe this landmark (see Figure 1(b) for a descriptor update given three sensors). The importance weight p for each bin is calculated according to a sigmoid function.

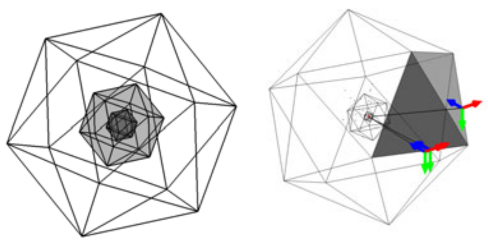

Figure 1: (a) Uninitialized HOC-descriptor around a feature point. (b) Update procedure with three sensor positions. Darker colors indicate a higher weighted bin.

The proposed descriptor allows us to add the following information to the map:

- spatially constrained visibility by increasing n,
- probable occlusions/dynamics by decreasing n,
- probably vanished features by looking at the histogram maximum.

3 Simultaneous Localization and Mapping

Simultanous localization and mapping is done with an incremental Structure from Motion framework estimating a 6DOF camera pose and 3D landmarks. For feature extraction and data association we make use of SIFT-descriptors while loop closure detection is handled through a vocabulary tree.

Data association and camera pose estimation is enhanced by selecting only those map points with an importance weight exceeding a predefined threshold. HOC-descriptors which produce a positive match are upweighted with the estimated sensor pose, whilst others are downweighted. To keep a constant map size we apply thresholding to the maximum of all bins of a descriptor.

4 Experimental Results

We performed a series of synthetic and real-world experiments, where several objects have been moved, deleted, occluded or re-appear after some time (see Figure 2). We compared the performance of the standard SLAM algorithm with its extension using the HOC descriptor by evaluating the map growth over time and the pose estimation error, where groundtruth was available. In addition we also evaluated re-localization performance (see Table 1). We greatly enhanced localization accuracy especially in high dynamic scenes and reduced map size enormously.

Figure 2: Real world experiment covering an office scene with various moved or deleted objects.

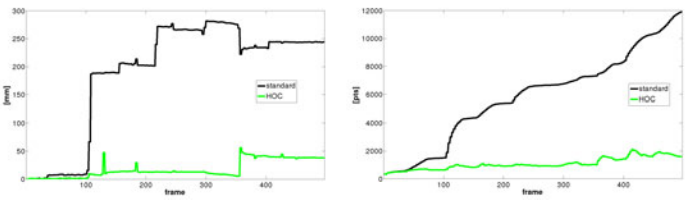

Figure 3: Pose Estimation accuracy and evolution of the map size over time for a static camera experiment.

testcase	TE [mm]		RE [deg]		map size [pts]	
	std.	HOC	std.	HOC	std.	HOC
synth. experiment	12.2	1.2	0.15	0.06	1024	295
static camera	30.7	2.5	1.51	0.11	6192	855
moving camera	X	X	X	X	15120	3680
relocalization	9.2	3.5	0.56	0.19	X	X

Table 1: Selected subset of various experiments. We compared translational (TE), rotational (RE) pose error and final map size between the standard (std) and extended HOC approach.

[1] A.J. Davison, I.D. Reid, N.D. Molton, and O. Stasse. MonoSLAM: Real-time single camera SLAM. *PAMI*, 29(6):1052–1067, 2007.

[2] E. Eade and T. Drummond. Monocular SLAM as a graph of coalesced observations. In *Proc. 11th IEEE Int. Conf. on Computer Vision*, pages 1–8, 2007.

[3] G. Klein and D. Murray. Parallel tracking and mapping for small AR workspaces. In *6th International Symposium on Mixed and Augmented Reality*, pages 225–234, 2007.