The paper presents a video matching approach to recognize complex human activities from realistic videos. The matching is done by a spatio-temporal relation match. The matching indicates how much two features are like and how close they are close to each other. A pairwise feature, is defined by its location in 3-D coordinates and the appearance in the local video. Temporal information indicates the position of a feature in video frames and time sequence. Because of that it is possible to detect complex activities even with partial occlusion. The spatio-temporal match is done by histogram match kernel. Each bin in the histogram combines two features when they are satisfied with some relation. A set of feature will be extracted from training data and be compared with ones in test data. The similarity between features is calculated by the histogram-based kernel and partial matching is also permitted. The features in the test data will vote for the most possible positions by comparing with the feature extracted in the test data. The location of features are normalized by the relative distance. By both localization and activity detection, multiple human actions can be detected in complex real videos.