With the exponential growth of the digital data, automated video content analysis has been drawing increasing attention of computer vision researchers. Effective modeling of objects, scenes and actions is critical for video content understanding. Recently, there has been a growing interest in deep learning, which has shown impressive results in speech and object recognition. In this dissertation, we propose several formulations and extensions of deep learning methods which learn hierarchical representations for video analysis, including complex event recognition, object detection in videos and measuring action similarity. For complex event recognition, we propose a novel unsupervised approach to discover data-driven concepts from multi-modality signals (audio, scene and motion) to describe high level semantics of videos. Our methods consists of two main components: we first learn the low-level features separately from three modalities. Then we discover the data-driven concepts based on the statistics of learned features mapped to a low dimensional space using deep belief nets (DBNs). For improving generic object detector in videos, we present a new model that learns the hierarchical object representations in a semi-supervised manner. It differs from the existing unsupervised feature learning methods in two ways: first it optimizes for both discriminative and generative properties of the features simultaneously, which gives our features better discriminative ability; second, our learned features are more compact, while the unsupervised feature learning methods usually learn a redundant set of over-complete features. For measuring action similarity, we describe a novel approach that learns the features and metrics directly from the data. We propose a generative plus discriminative learning method based on gated auto encoders to simultaneously learn the features and their associated metrics. Extensive experiments with quantitative and qualitative results on the three tasks demonstrate the superiority of our proposed models.