With the growing importance and volume of multimedia data, interactive search systems are essential to help users efficiently search for specific video sequences based on content. One common task is known scene retrieval, where users try to find a particular scene in a large collection of videos. However, the description of a known scene can be subjective, influenced by the perception and experience of individual users, as well as the differences between human and machine perception. In this paper, the effectiveness of an interactive retrieval system combined with image classification generated by a CLIP neural network is investigated to address this problem. V3C datasets and Marine Video Kit are used to verify the effectiveness of the proposed method. Software is also presented that allows data collection for experiments and subsequent evaluation using a web interface.