Problem Identification

I want to write some software that can convert a video stream into sound. The sound should vary according to the distances to the objects in the video.

An assumption that I will be making is that the objects in the environment being videoed are not moving, only the camera. This allows us to use the fact that the observed motion of an object as a result of a change in perspective in greater for objects that are closer to the observer.