Alpha based similarity detection

* Since water is naturally transparent/ semi-transparent, using alpha transparency as a water quality indicator
* Pure water is transparent, (has a low alpha value)
* Murkey dirty water appears opaque, (has high alpha value)
* Matter in the water would increase opacity

Pit falls

* In a perfect world this would be an ideal method, however a fundamental issue in image processing is most standard images are saved as opaque (with alpha set to 255) by default even if there is transparency
* What looks transparent to our eye as humans is actually just light-coloured pixels (light blue, and white) and true digital transparency requires and alpha channel
* Alpha channels are usually added with manually with image editing software, and normal cameras do not capture alpha data
* To overcome this, we will make a function to simulate transparency by converting HSB (Hue Saturation and Brightness) colour space, which colour model often used in graphics software for selecting and manipulating colours. It's a more intuitive way for humans to describe colours compared to RGB
* Converting to HSB separates colour information into perceptual components that align better with how humans (and nature) distinguish water characteristics.

More cons:

**1. False Negatives (Missing Polluted Water)**

**Problem**: Strict HSB thresholds might exclude:

* **Murky brown water** (hue shifts toward yellow/brown, ~0.1-0.2)
* **Oil slicks** (low saturation, high brightness, S < 0.2, B > 0.9)
* **Turbid water** (high brightness due to sediment, B > 0.8)

**Example**: A polluted river might be misclassified as "non-water" if its hue falls outside 0.45–0.65.

**2. Edge Effects at Water-Land Boundaries**

* **Problem**: Pixels at the shoreline might be:
* **Misclassified**: Mixed water/land pixels (e.g., shallow water with visible sediment).
* **Omitted**: Critical for detecting erosion or runoff pollution.

These might cause GED to underestimate changes near coasts.

**3. Dependency on Lighting Conditions**

* **Problem**: HSB values shift under:
* **Sun glare** (inflates brightness, B → 1.0)
* **Cloud cover** (reduces brightness, B → 0.5)

Causing the Same water to be classified differently across images.

* + 1. **Static Thresholds vs. Dynamic Water Conditions**

**Problem**: Fixed HSB ranges fail for:

* **Glacial lakes** (turquoise hue, H ≈ 0.5–0.6).
* **Tropical waters** (higher green, H ≈ 0.35–0.45).

**Shadows**: Can mimic low saturation (add brightness checks).

**Sun Glare**: Appears as foam (use gradient analysis to distinguish).

**Shallow Water**: May show bottom sediments falsely as pollution.

Because of these factors we have decided to focus only on potential water pixels by weighing their GED costs higher when computing GED for water quality analysis. This will increase the performance as unnecessary GED calculation will be affect water analysis less, however at the cost of less actual water pixel regions. In this way we feel we balance the pros and the cons.