TRACKING STRATOCUMULUS CLOUD BREAKUP WITH IMAGE SEGMENTATION

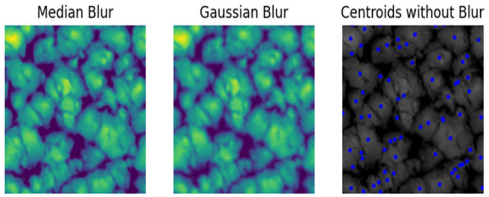
SAM SPIEKER

INTRODUCTION

STRATOCUMULUS CLOUDS PLAY AN IMPORTANT ROLE IN PREVENTING CLIMATE CHANGE. THESE LOW-LEVEL (1,000-6,500 FT) CLOUDS ARE THICK AND COVER LARGE AREAS OF THE EARTH. THEY REFLECT SOLAR RADIATION FROM THE SUN, WHICH HELPS COOL THE EARTH'S SURFACE. HOWEVER, THERE IS BELIEVED TO BE A DECREASE IN STRATOCUMULUS CLOUDS. INCREASING GREENHOUSE GAS EMISSIONS TRAP MORE HEAT AND BREAKUP STRATOCUMULUS CLOUD DECKS. THIS BREAKUP CAUSES LESS CLOUD COVERAGE, THUS PROMOTING GLOBAL WARMING. IT'S A POSITIVE FEEDBACK LOOP THAT FURTHERS GLOBAL WARMING AND STRATOCUMULUS CLOUD BREAKUP. WE CURRENTLY DO NOT HAVE AN ACCURATE PERCENTAGE OF CLOUD COVER NOR AN UNDERSTANDING OF HOW THAT PERCENTAGE HAS CHANGED OVER TIME.

METHODOLOGY

THE FIRST STEP REQUIRES THE USE OF A BLUR METHOD. BOTH A GAUSSIAN BLUR AND A MEDIAN BLUR WERE TESTED FOR THIS. THEY REQUIRE A SPECIFIC KERNEL SIZE TO DETERMINE THE AREA THAT IS BLURRED, I.E. A LARGER KERNEL RESULTS IN A STRONGER BLUR AND A SMALLER KERNEL RESULTS IN A WEAKER BLUR. FURTHER, THE KERNEL SIZE MUST BE AN ODD NUMBER IN ORDER FOR A CENTRAL PIXEL TO ALWAYS BE PRESENT. VALID KERNEL SIZES ARE 3X3, 5X5, 7X7, AND SO ON. THE MEDIAN BLUR REQUIRES A SQUARE IMAGE AND ENSURES THAT A SQUARE KERNEL IS CHOSEN. THE GAUSSIAN BLUR, HOWEVER, DOES NOT REQUIRE A SQUARE IMAGE, NOR A SQUARE KERNEL. SINCE THE DATASET IS ONLY SQUARE IMAGES,



ONE OF THE ALGORITHMS THAT WILL BE ANALYZED IS K-MEANS CLUSTERING. THIS PROCESS INVOLVES 4 STEPS:

1. K DATA POINTS ARE RANDOMLY SELECTED AS INITIAL CENTROIDS.

2. EACH PIXEL IS ASSIGNED TO A CENTROID BASED ON A DISTANCE METRIC.

3. THE AVERAGE IS TAKEN OF EACH CLUSTER AND ASSIGNED AS THAT GROUP'S NEW CENTROID.

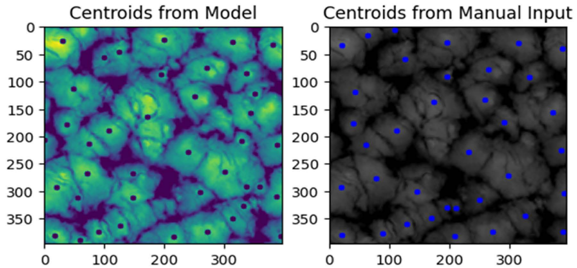
4. STEPS 2 AND 3 ARE REPEATED UNTIL THERE ISN'T A SIGNIFICANT IMPROVEMENT.

K-MEANS WORKS DIFFERENTLY WHEN USED FOR IMAGES RATHER THAN GRAPHED DATA POINTS. IT DOES FOLLOW THE SAME PROCESS OF REPEATEDLY TAKING THE AVERAGE OF EACH CLUSTER, BUT WITH PIXELS. FOR EXAMPLE, IF K=4, 4 RANDOM PIXELS WILL BE ASSIGNED AS CENTROIDS, THEN ALL OTHER PIXELS WILL BE ASSIGNED TO ONE OF THE CLUSTERS BASED ON THE SIMILARITY OF PIXELS. THE SIMILARITY IS ASSESSED BY THE EUCLIDEAN DISTANCE OF THE RGB VALUES. NEXT, THE AVERAGE IS CALCULATED OF EACH CLUSTER AND REASSIGNED AS THE NEW CENTROID. THIS PROCESS REPEATS UNTIL NO SIGNIFICANT IMPROVEMENT OCCURS.

THE THRESHOLDING METHODS THAT WERE TESTED ARE A SIMPLE THRESHOLD, OTSU, GAUSSIAN ADAPTIVE, AND MEAN ADAPTIVE THRESHOLDING. BOTH SIMPLE AND OTSU THRESHOLDS ARE PERFORMED GLOBALLY. A SIMPLE THRESHOLD ALLOWS YOU TO CHOOSE A VALUE AND ANY VALUE HIGHER THAN THE THRESHOLD IS SET TO 1, USUALLY LABELED BY WHITE. THE VALUES BELOW THE THRESHOLD ARE SET TO 0, USUALLY LABELED BY BLACK. THE OTSU THRESHOLD WORKS THE SAME, BUT IT WILL AUTOMATICALLY CHOOSE A THRESHOLD BASED ON THE IMAGE DATA.

RESULTS

THE FINAL MODEL USES A MEDIAN BLUR WITH A 9X9 KERNEL, K-MEANS SEGMENTATION WITH K=6 CLUSTERS, THEN A SIMPLE THRESHOLD, WATERSHED METHOD AND FINALLY A CENTROID CALCULATION METHOD. THIS MODEL HANDLES THE EARLIER ISSUE VERY WELL; LARGE CLOUD NETWORKS ARE BROKEN UP INTO EACH INDIVIDUAL CLOUD AND THE CENTROIDS ARE VERY ACCURATE. THIS MODEL ACHIEVED A FINAL AVERAGE DISTANCE BETWEEN MATCHED CENTROIDS OF 15 PIXELS, ONLY 2 UNMATCHED CHOSEN CENTROIDS AND 8 UNMATCHED MODEL CENTROIDS.



CONCLUSIONS

BY USING IMAGE SEGMENTATION AND THRESHOLDING METHODS, WE HAVE DEVELOPED A MODEL CAPABLE OF DETECTING AND TRACKING CHANGES IN CLOUDS OVER TIME. THE FINAL MODES, UTILIZING A MEDIAN BLUR, K-MEANS SEGMENTATION, SIMPLE THRESHOLD, AND WATERSHED METHOD, HAS DEMONSTRATED SIGNIFICANT IMPROVEMENTS IN ACCURATELY IDENTIFYING CLOUD CENTROIDS, DESPITE THE ABSENCE OF GROUND-TRUTH MASKS.

OUR FINDINGS UNDERSCORE THE IMPORTANCE OF ACCURATE CLOUD COVER ESTIMATION FOR BOTH KNOWLEDGE OF CLIMATE IMPACT AND WEATHER PREDICTION. THE DEVELOPED METHODS PROVIDE A START FOR FUTURE RESEARCH IN CLOUD DETECTION AND SEGMENTATION. FURTHER IMPROVEMENTS OF MORE COMPREHENSIVE DATASETS COULD REFINE THESE