

Hue-Based Parallel Image Recognition

Birch Sztabnik, Jerry Lei, and Jonathan Cheng



Abstract—In this paper, we discuss our implementation of a massively parallel image recognition program using hue-based hashing algorithm. The process of how hashing works when hue is a factor and how to parallelize it is described. Also discussed is the results of such a system's performance in terms of detection and speed.

1 INTRODUCTION

A common technique amongst other image hashing tools is to resize the image down to an extremely small size before hashing. This removes any details that could cause differences between images. This technique is often paired with reducing color to grayscale to avoid color differences [4]. A side effect of this process is that images can not be searched for accurately since all detail is removed. The purpose of our test was to design an parallelized algorithm that would function on original size, color images and to determine its effectiveness.

2 HYPOTHESIS

We believe that we will get an increase in performance with more nodes, but the increase in benefits will stagnate. Our dependency on message passing among ranks is fairly limited, and the core of our wait time depends on the amount of computation. As we increase the number of ranks, the work is distributed to different ranks, and each rank would be responsible for less. However, the sizes of search that we need to compute stay constant, and the largest search job for a single image cannot be split into multiple ranks. Once we increase the number of ranks to the point where the largest job is the only job the rank has to calculate, we do not suspect there will be any more speedup.

3 PROGRAM IMPLEMENTATION

3.1 General Image Functions

Some image manipulation functions are used. Most commonly used are the image resizing functions. One version resizes by percent. This version is used in the workload algorithm for creating variations for each rank to hash. The other version resizes to predefined dimensions. This is used prior to distributing hashing work in order to make certain that the number of hashes that can be made is an integer value. Both operate on integer division to fill the resized image.

A helpful function to use before starting any hashing is to crop the board. The ideal search image to be used is one on a solid white background. Then it can be cropped by the program to reduce as much background as possible. This is an important step for the search image. The fact that white pixel are ignored in hashing and that image size is larger than it need to be means that the scoring distribution would be skewed. Images with large white borders would report zero for pixels on the edges, which make it difficult to score effectively.



Fig. 1. The scoring system correctly matched the PDA using a stock photo. Distribution is important in determining between extremely similar objects. The white blocks represent hashes.

3.2 Hashing Function

The image searching algorithm depends on the hashing of the image by color and gradient. Hashing requires a 9x8 block of pixels. The function iterates over the block, converting the current pixel and its right neighbor to HSV color.

HSV is required for useful comparison as small variations in RGB color components will drastically change the color. Initial tests were done using RGB and the worker function would determine that the hash had no match, unless the image was a direct cutout. HSV also enables easy identification of white backgrounds on images.

The most important variables are the gradient hashes for hue, saturation and value. The three are calculated in the same fashion. The pixel on the left is compared to its neighbor to the right on the x-axis. If the hue, saturation, or value is less than its neighbor, the bit is set to 1 or 0 otherwise. Hashes are stored in a 64 bit integer. This allows us to write each inequality to each bit. There are only 64 comparisons made, as the 9th column is used for comparing to the 8th column.

There are also three other metrics stored. One is the average hue of the 8x8 block. The other is the normalized sum of the 8x8 block hues, which is known as the secondary hash. These two calculations are used in the worker function. The top left corner pixel information is also stored and will be used later in the filtering function to determine if the hash is to be skipped.

3.3 Hashing Allocator

The hashing function forms the first half of our recognition algorithm. Two parts are needed for operation; hashing of the original image and the hashing of the search image. Both use the same fundamental hashing function, but the application is different. Prior to hashing, the search image is resized to be divisible by 8 across the y-axis and divisible by 8 plus an extra pixel on the x-axis. This is because hashes are 8x8 pixels in area and require information on the right neighboring pixel.

On the original image, a hashing function is run on every pixel position in the image. This is because the location of the search image is not known, so every position must be considered. That property removes the need for resizing for divisibility. On the search image, hashes are generated for every 8x8 square.

In order to track the amount of matches, an array called the hitbox is made. This is the same size as the original image, minus the edges cut off due to the hash size. To parallelize hashing each image, pthreads are used with a defined maximum number of threads.

A problem we encountered was distributing work among threads without getting overlapping hashes. Methods using modulus and division were not effective due to overlap and high amounts of computation. Some initial attempts were dividing the image into smaller blocks until less than 64 threads could be used. These blocks would be comprised of smaller 8x8 blocks that a thread could iterate over and hash. Guaranteeing that the image will not have overlap or missed pixels is difficult if there the dimensions have few factors of 8. This method usually allocated work for a less than optimal number of threads. In the case of a 200x304 image, the maximum amount of threads that could be created would 50 with a super-block size of 8x152 per thread.



Fig. 2. Threads hop over each other's block to ensure no overlap.

Our solution was to have staggered starting points for the threads. The rank calculates the maximum number of threads that can be generated which is defined as the area of the image divided by the area of the hash. It is almost always possible to generate 63 threads + 1 rank call. The rank begins creating threads at 8 pixel in on the x axis, and creating a thread every 8 pixels, wrapping around to the next column if it extends past the bounds. This also means older threads have more work. After creating all threads, the rank calls the same function to calculate, starting at position 0x0, for a total of 64 calls on the hash worker function. When the rank finishes hashing, it joins with the threads it created.

After all threads are joined, the rank decides the optimal distribution of hashes using the search image and calls the to match that distribution on the filled hitbox. Implementation details are discussed in the Scoring System section.

3.4 Hashing Worker

Several operations happen in the thread worker function. This is where the hitbox gets filled in. The worker begins by hashing an 9x8 block at the position the rank gave it. To avoid overlap rehashing pixels since the threads begin adjacent to each other, the thread must increment by 8 times the total number of workers. This ensures no rehashing of pixels occurs. Of course, wrap around must exist. Attention must be given to the fact that incrementing the x position by this amount can put it out of bounds several times through. That means the x position must be reduced on a loop until it can reach somewhere on the image.

There are a few metrics that the hash function gives us. First, the top left corner pixel of the hash is passed through a filter to determine if the pixel is near white. If it is, the hash is ignored. We chose to filter like this because our search images were placed on white background and white was to be treated as transparency. It makes the distribution array generated from the hitbox more accurate. If the hash is valid, the worker must traverse the hashed original image, comparing between.

First, the hamming distance is calculated between the search image and original image's hue, saturation, and value gradient hashes. The distances are weighted, and then summed together. We determined that weighting the hue distance the most gives good results. The worker also checks if the corner pixel hue is similar and that the combined difference of the saturation and value from the original image corner pixel are not too different. Not including this test initially gave problems where dark pixel blocks were matching with light pixel blocks since the hues were similar despite the lightness being completely different.

Also checked is if the secondary hash and the average hue of the search image is close to the original image. A version of the worker function retained the best secondary hash and average numbers. The resulting hitbox was scarcely populated and did not match with the best match possible.

If the hash passes, the best weighted sum is retained to see if any other positions in the hashed original image match more closely. When the whole board is done being searched, the hitbox is incremented at the position of the best found match in the original image. This hitbox is shared amongst the threads and rank, so it is mutex protected.

3.5 Scoring System

The goal of the scoring system was to be able to determine which area in the hitbox most closely matched the search image. Clusters of hits indicate a match, but random noise must be avoided. We experimented with multiple mechanisms to use as a scoring system. We started off by calculating each hits distance to the closest hit by propagating outward in a spiral pattern. We used this distance to check how close points are to one another, and gave a search box a higher score if the spread was tighter. This was not the best metric because random noise influenced the score significantly, and caused inconsistent results when searching using a wide range of scales.

We also attempted to average the location of all successful hits and determine how far that was away from the center of the current search image to measure how centralized all of the points were. This was a satisfactory metric; however searching through the entire original image would allow for edge cases that receive good scores. For example, a scoring chunk with a random hit on each of the four corners would receive a similar score to a cluster with 500 hits in the center of the search chunk.

In order to remedy this, we attempted to make a distribution by analyzing a search chunk twice. We would base a score off of the entire search chunk, then segment it into 4 equally sized chunks and analyze their scores. We were under the assumption that a search chunk with a high score and smaller chunks with scores that do not have a lot of variance would likely have the image in it. We could not determine a scoring system that would allow for this to work. All of these methods did not work perfectly, but they began taking the direction of a scoring system based off of distribution and density.

x_0	x_1	x_2	x_3	x_4	x_5
0.1092	0.1533	0.1983	0.2195	0.1890	0.1306

y_0	y_1	y_2	y_3	y_4	y_5
0.0368	0.1112	0.2152	0.2383	0.2226	0.1758

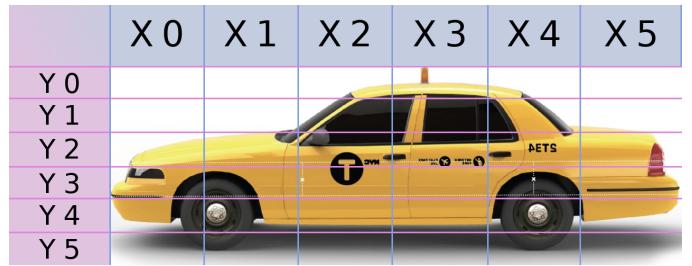


Fig. 3. Distribution breakdown for taxi example.

We computed distribution by dividing up the search area and the original search image into equally sized buckets both lengthwise and widthwise, representing two one-dimensional arrays. For the search image we filled up the buckets with the number of pixels that passed filtering in its section, and divided it by the total number of pixels that passed filtering in the original search image. Essentially, the pixels across the y-axis for the width of each bucket are compressed into a one-dimensional array that represents the distribution of the pixels across the x-axis. We calculated the distribution to be used during each scoring based using the same buckets, but used the hitbox received from the hashing process for a given search image. The distribution measurement can be thought of as a pseudo-shape detector.

This was not enough to give us accurate results of image recognition so we also incorporated density of hits of each search area to the score. The calculation for density was simply the number of total hits in the search area divided by the total area of the search image. Scoring was largely dependent on the search image, so we are unable to determine a set score that could decide whether the search image exists inside the original image.

3.6 Scaling

Implementing a variety of scales to compute was an integral part in finding out where the search image was in the original image. We scaled based upon the search image size and the original image. Our logic was that the image that object we are looking for in the original image cannot be bigger than the size of what we are searching for. Therefore we set the upper bound for scaling the search image to match the size of the original image. We decided on 10% of the size of the search image as a lower bound. Our formulation of search sizes is as follows:

Let U = the upper bound of scales

Let L = the lower bound of scales

Let s = the number of scales inbetween U and L

$$\text{Distance between scales} = \frac{U - L}{s}$$

$$\text{Possible scales} = \frac{U - L}{s} \times i,$$

where $i = 0 \rightarrow s$

3.7 Workload balancing

We came across a problem where we had to distribute the number of search images, or jobs, into different ranks to do independent calculations. The algorithm distributes the workload of search images into ranks such that their calculation times are the same. We valued the workload of each job with the scale of the search image squared. This represented the area that the rank must search through. The number of ranks given to the program corresponds to the number of machines that we have to process the calculations. This is a load balancing problem. There does not exist a polynomial time algorithm to distribute the workload optimally, but a sorted greedy problem gives a good approximation. We ran the greedy algorithm by generating the jobs in descending order and assigning them to machine with the least current load. Every rank did this calculation because the result of the greedy algorithm is deterministic and it would be faster to do the calculation than to have one rank do the calculation and send it to the other ranks. This algorithm has a $\frac{3}{2}$ approximation factor, and our experiments showed that the load was balanced within 1% with more than 20 scales of images.

3.8 MPI Implementation

Our initial idea was to combine all of the hitboxes that every rank computed, but we recognized that because we are hashing many sizes, larger scales of base images would overpower the smaller scales, and the scoring system would have a difficult time determining where the image truly was. Also, passing our hitboxes across ranks would too take too long and summations of these large arrays may take too long.

Instead, we implemented a scoring system such that if we were given any two hitboxes, and two search sizes, we can determine which was better. We assigned a score value to each search size and reduced the best score into a single rank. That rank reported the score and overlaid the resulting bounding box over the original image to display our result.

The final implementation has ranks using an MPI_Allreduce in MPI_MAXLOC mode. The maximum score that each rank computes is shared with its rank number. The rank with the highest score draws a bounding box on the original image and writes it to a file.

4 RESULTS

4.1 Matching Results

Image matching depends on a wide range of factors. Using search images that have too much glare or are too shiny will have most of their information filtered or will not match well. Colors such as gray also pose trouble since two similar looking grays can have very different hues.

The test chosen was to search for a button on a coat in good lighting condition. The search image appears twice in the image, with four more similarly colored sections further away. The algorithm would have to determine the newspaper photos did not match and match the circular search image distribution with the button, not the photo. The button was copied from the original image and placed on a white background. It was then searched on the original image. The button was 4 pixels away from its actual position in the image. This represents the best-effort scenario in which the perfect match conditions are set. This is used as the control for this test. A different photo of the same button was used as a search image.



Fig. 4. The button cut out from the original image (left). A different photo of the same button which was used as the search image (right). The perspective is vastly different from the original photo.



Fig. 5. Multiple areas of same colors will cause mismatches if search set is small.

The scales tested were powers of two from 2 to 64. Pixel distances are from the control to the match and area is rated in terms of the cutout button. Distances are on a logarithmic scale.

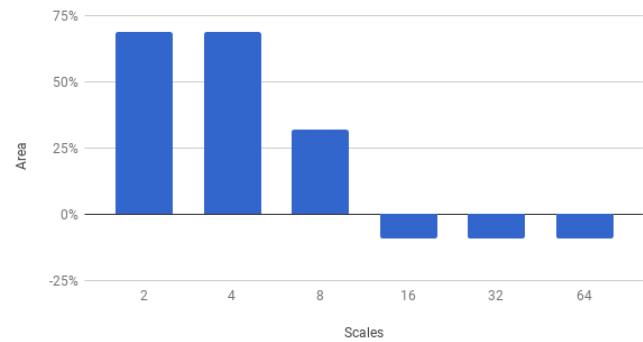


Fig. 6. Area of match to best-effort vs. scales used

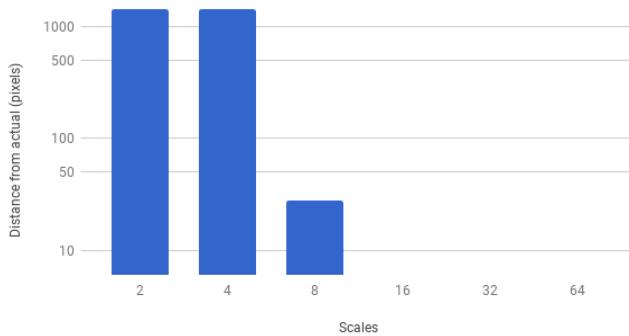


Fig. 7. Distance of match to best effort vs. scales used

Using two scales is usually never enough to determine a close match in an original image, unless the workload so happens to chose the correct scale the first time or the search image is a direct cut-out from the original. The two and four scale tests faired poorly. In comparison to the area of the best match from 64 tests, the area was 47% larger at best. Four scales had the same result but a closer area was generated at 9% larger than the best of the 64 scale test. 8 scales is where improvement was seen with a smaller over-approximation. The area was 32% larger than the best-effort test, but would be acceptable if it were to be cropped. 16 scales and above produced the best results. This is when the workload calculated the ideal size and the scoring system was then able to identify the correct cluster. The area matched was 9% smaller than the best effort in 16 scales and up. Past 16 scales, a better fit would not be calculated. The next best score determined in the rank comparison process was 20% lower. It is possible that generating beyond 64 scales would yield better results, but the improvement would be minimal.



Fig. 8. Left to right: 2,4 scales; 8 scales; 16,32,64 scales.

For two and four scales, the algorithm failed to match the button. It instead matched with the blue photos in the newspaper, which was to be expected because of poor scoring due to poor scale sizes. The 8 scales test was able to land on the button at only 28 pixels away from the best-effort. Tests with greater than 16 scales showed excellent results being nearly 6 times closer than the 8 scale test. In some cases, the search image cannot get closer to the best-effort with more scales. Disregarding image differences, this is because the image is resize for divisibility. The search image match in the 16 scale test of size 153x152 has an automatic distance of one away from the real position since it was not completely divisible by 8.

Another thing to be noted is that the right button was not matched. This is because of the reflection down the middle. These pixel would be filtered out during hashing and alter the distribution. A circle has the highest distribution in the center buckets, which filtering would remove. This means that the search image would never match with the right button.

4.2 Performance Results

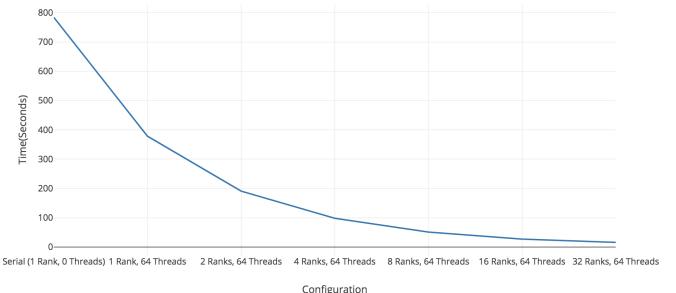


Fig. 9. Compute time vs. ranks used at a constant workload

We experienced a significant amount of scaling with an increased number of nodes. We achieved double the performance whenever we doubled the nodes. In our testing, we used the maximum number of nodes possible per node and varied the number of ranks. The workload was kept constant at 64 scales. The sizes of these scales were kept constant through workload distributions. Original image size was held constant. Code was compiled using mpixlc_r and optimization level 3.

Initial tests were ran using a single pthread mutex to protect the shared hitbox. The slowdown from this was not made apparent until more ranks were used.

We also checked how much overhead we have from using a single mutex rather than having an array of mutex, and found that there is a significant amount of slowdown with an increased number of threads.

We tested with a single node, using a varying number of ranks and threads that multiplied up to the maximum 64 per node. We noticed a significant amount of overhead was resulted from the using a single mutex lock for the entire hitbox among all threads. When a search starts, all the threads are active and will try to lock at the same time which is inefficient. This problem lessens as threads complete operation.

As a result, we created a board of mutexes to that gave access to individual indices in the hitbox. This significantly reduced our computation time. After making this change, we reran the same test and found no differences in compute times with a constant number of threads.

Further tests were also done to determine the usefulness of pthreads in relation to the number of ranks. The number of ranks plus the number of threads was kept equal to 64. There was little slowdown versus using the maximum number of threads possible. This could be due to the gains received from further parallelizing the hashing process being passed by slowdown from thread creation. Slowdown could

also be due to spinlocking even though a 2D mutex array is in use. Areas where the match is apparent early on will get hit frequently, at a scale of hundreds of times. This delays hashing for a long period of time as the thread could be waiting in a line. More threads make this problem worse. Larger images do not benefit from less coinciding hits.

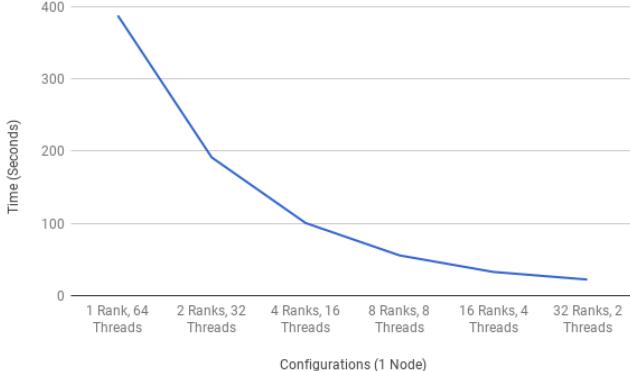


Fig. 10. Keeping a constant number of workers.

We also analyzed the distribution of workload between the ranks and found that the average of standard deviations among all ranks was 1.44 seconds of compute time.

5 FUTURE WORK

Given more time, we would attempt to make the hashing algorithm on the hitbox more position-aware, meaning that it will not consider a set of pixels a hit if it is too far away from an already discovered cluster on another rank or a previously considered hash. Further improvement could be made by having hashes from the search image remember their position relative to their neighbors. This would help to maintain a cohesive image rather than a wide spread over the hitbox. This method would require high levels of communication between the threads which would slow performance but could yield more accurate results. With a more meaningful scoring system, we could also attempt to share information about what jobs to skip. Meaning, if we were to get a consistently decreasing score while lowering the scale, we would skip scales that are lower than that. We could do this by leaving a thread open with the sole purpose of sending/receiving the current status of scores and scales that other ranks have been computing.

We would also like to be able to come up with a better way to filter our images to reduce noise beyond filtering light colors. The current iteration of the filtering mechanism is used on hashing and distribution detection which greatly influences image detection. Currently, it aggressively filters out near-white pixels and grays. This means it has difficulty detecting shiny objects or object light in color. Ideally it would have different functionality based on what color is passed in to it and have some sort of awareness to what is background on a cropped image.

Noise from cameras can be a problem on low light or heavily zoomed images. As image size grows, the problem

becomes worse since the program will start hashing noise. A noise detection and reduction algorithm would have to be implemented. Effects of noise can be reduced by shrinking the image down.

There were also functions for image skewing and rotation created, but were not used due to long runtimes [2]. The workload algorithm and hashing does repetitive work which slows operation. Introducing rotation and skew to the workload would drastically increase runtime. The current hashing function handles perspective shift well but cannot account for rotation. Instead of hashing every pixel, most other image recognition algorithms rely on a random sampling of pixels to check [1, 3], and we would like to do something similar given more time. We would also like to be able to check for better combinations of threads and ranks to reach greater parallelism and speedup.

6 SUMMARY

Our work proposes an algorithm that can take advantage of parallelism in order to search for images at original scale in color. Our hashing algorithm is comprised of complimentary components which aid in producing a meaningful hitbox. The scoring system takes image composition into consideration in order to provide an accurate match. This is combined with a parallelized system that distributes work and further parallelizes it using threads

Our results show that under photos with good lighting condition, our program can produce results that are as accurate as 9% difference in size and within the distance of one hash to the optimal. This can be done using search images that have much different perspective location than the original and have minor color differences. The workload is done faster as the number of ranks increases. In most cases, match accuracy increases as the scale set increases. We feel that hue-based image recognition is a viable method of image search.

7 DIVISION OF LABOR

Birch: Came up with hashing algorithm, developed general image functions, brainstormed ideas for scoring system, testing and setting up on BG/Q

Jerry: Implemented scoring system, assisted with hashing algorithm and general image functions, balanced workload for threads & ranks

Jonathan: Ran tests for the program, generated results for graphs, brainstormed ideas for scoring system.

REFERENCES

- [1] M. Guan, C. Wen, K. Y. Lim, M. Shan, P. Tan, C. L. Ng, and Y. Zou. Visual tracking via random partition image hashing. In *2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pages 1–6, Nov 2016.
- [2] A. W. Paeth. Graphics gems. chapter A Fast Algorithm for General Raster Rotation, pages 179–195. Academic Press Professional, Inc., San Diego, CA, USA, 1990.
- [3] A. Swaminathan, Y. Mao, and M. Wu. Robust and secure image hashing. *IEEE Transactions on Information Forensics and Security*, 1(2):215–230, June 2006.
- [4] B. Yang, F. Gu, and X. Niu. Block mean value based image perceptual hashing. In *IIIH-MSP*, pages 167–172. IEEE Computer Society, 2006.