Region-based Querying of Solar Data Using Descriptor Signatures

Juan M. Banda, Chang Liu
Department of Computer Science
Montana State University
Bozeman, MT, 59717-3880
juan.banda, chang.liu@cs.montana.edu

Rafal A. Angryk
Department of Computer Science
Georgia State University
Atlanta, GA 30302-3994
angryk@cs.gsu.edu

***Abstract*—In this work we present our first results on the ambitious task of providing region-based querying capabilities to our existing Solar Dynamics Observatory (SDO) content-based image-retrieval (CBIR) system. By taking advantage of pre-computed image descriptors, we calculate region-based histogram signatures for our training set of previously identified solar events. With these signatures we then explore the possibility of retrieving new similar solar events solely based on these scale and rotation invariant signatures. In this work we present our proposed methodology, initial retrieval results, verification and validation mechanisms, and propose future work with possible applications for other fields in astronomy.**

***Keywords*—CBIR, region-based querying, image retrieval, image classification**

I. Introduction

With the launch of NASA's Solar Dynamics Observatory (SDO) mission, a new era of large-scale data mining in the field of solar physics has started. With over 70,000 images being sent to Earth on a daily basis, traditional computer-based methods of analyzing solar data have become obsolete and newer large-scale big data-driven technologies have to be developed to cope with new data deluge. While individual systems have been developed to help identify particular solar events (flares, filaments, sigmoids, active regions, etc.) in the SDO repository as indicated in [1], we have been developing a more generalized event finding system. Called the SDO Content-Based Image Retrieval (CBIR) system, it utilizes machine learning and information retrieval methods to provide an all-purpose tool capable of identifying multiple solar events solely on their visual characteristics. The system's findings and results can be found in [2]–[6] and a currently working version of said system is available at **http://cbsir.cs.montana.edu/sdocbir**.

This system currently searches for similar images in a full-disk manner; however, this work outlines our preliminary steps and results on integrating region-of-interest based querying (outlined in Figure 1) to provide researchers with a higher fidelity search of particular areas in solar images. Our region-based querying analysis utilizes pre-existing solar event labels taken from Heliophysics Event Knowledge (HEK) base (and its respective Feature Finding Team (FFT) modules) to build standard image signatures for each event. Making said image signatures scale-and-rotation invariant will allow us to query the extensive SDO dataset using images or image segments of any size and covering any location in the sun. It also enables the use of images from non-SDO catalogs as as queries to search the SDO repository for similar events.

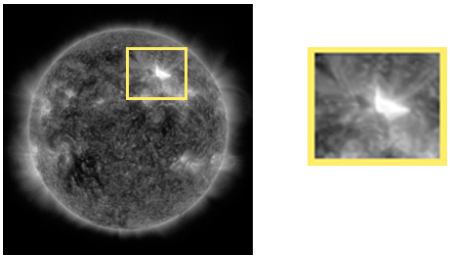

Fig. 1. Sample region of interest selection that includes a Flare (FL) solar event. Image signature is calculated based on the selected region only

In this work we introduce our testing dataset and mechanisms for generating our image signatures; we also showcase multiple retrieval scenarios to verify the effectiveness of our image signatures. We then provide validation for our results using machine learning classification algorithms on our image signature data. The remainder of this paper is structured as follows: Section II provides a brief background introduction to our work. Section III describes the dataset we are currently using for this work. In Section IV we detail the methodology behind our experiments. We then present the results of the experimental evaluation in Section V. Finally, in Section VI we will highlight our general conclusions, and in Section VII we conclude the paper with future work and other potential research applications.

II. Background

With the rapid growth of large image databases, querying in an effective manner has become more critical. Currently some of the most popular image query engines rely on the comparison of meta-data or textual tags associated with the images [7]. A standard practice in solar physics was to rely on experts to tag images—a task that is unfeasible now.

Over the years, a number of content-based image retrieval (CBIR) systems and image retrieval algorithms have facilitated these general purpose querying tasks, but they quickly falling behind thanks to the data deluge that scientific instruments generate (i.e. in solar physics and astronomy). Systems such as *Photobook* from MIT allow users to conduct the retrieval of images based on several different image features. This system works by reducing images to a small set of perceptually-significant coefficients [8] and then calculating similarities.

Other systems based on low-level image features include *QBIC* from IBM [9], *Candid* [10] and *Chabot* [11]. Most image retrieval techniques rely on attributes such as shape, color, or texture in a full-image manner, which is something we expand upon with our current approach. *Blobworld*, developed by University of California, Berkeley, is an example of an image index and retrieval system based on finding coherent image regions that are related to objects [12]. In the literature, histogram based query methods are often utilized to facilitate the feature comparison among pixels (and descriptors) in the region of interest. This approach is motivated by the importance of the interrelation among pixels in particular regions of interest. The main benefit from this approach is that the descriptors generated are invariant under rotation, translation and scaling, allowing us to compare against regions of interest from different format images and images scaled to bigger or smaller sizes.

III. Dataset

For this work, we used solar images from the *Atmospheric Imaging Assembly* (AIA) module of the *Solar Dynamics Observatory* mission (SDO). The AIA module captures eight high-resolution (4096-by-4096 pixels) images every 10 seconds—more details of this instrument are available in [1], [13]. In Schuh et al. [14], an SDO benchmark dataset was introduced. This dataset spans a six-month period of data and contains over 15,000 images in two wavebands featuring 24,000 event instances of six different event types. This image data is presented in a processed form where each image is segmented into a 64-by-64 grid of cells, and ten image parameters are extracted from each cell. In the work performed by Banda and Angryk [3], [4], many possible parameters were tested for solar images based on factors such as computational expense, and classification accuracy. The ten image parameters that we chose to characterize the solar images are: entropy, fractal dimension, the mean intensity, the third and fourth moments, relative smoothness, the standard deviation of the intensity, Tamura contrast, Tamura directionality and uniformity. These image parameters and corresponding formulas are listed in Table I.

TABLE I. Image Parameters.

Label	Image Parameter	Formula
P1	Entropy	$E = -\sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i)$
P2	Fractal Dimension	$D_0 = \lim_{\epsilon \to 0} \frac{logN(\epsilon)}{log\frac{1}{\epsilon}}$
P3	Mean	$m = \frac{1}{K} \sum_{j=1}^{K} z_j$
P4	3^{rd} Moment (Skewness)	$\mu_3 = \sum_{i=0}^{L-1} (z_i - m)^3 p(z_i)$
P5	4^{th} Moment (Kurtosis)	$\mu_4 = \sum_{i=0}^{L-1} (z_i - m)^4 p(z_i)$
P6	Relative Smoothness	$R = 1 - \frac{1}{1+\sigma^2}$
P7	Standard Deviation	$\sigma = \sqrt{\frac{1}{k} \sum_{j=1}^{k} (z_j - m)^2}$
P8	Tamura Contrast	*Tamura, Mori and Yamawaki [15]
P9	Tamura Directionality	*Tamura, Mori and Yamawaki [15]
P10	Uniformity	$U = \sum_{i=0}^{L-1} p^2(z_i)$

z represents an image cell, z_i is the i-th gray level, m is the mean, $p(z_i)$ denotes the gray scale histogram of the i-th gray level in the cell. z_j is the value of each pixel of the image, and j will go up to K where K is the number of pixels.

For the purposes of this paper, we chose a one month subset of the benchmark dataset from [14]. We selected three wavelengths and three different event types: Sigmoid (SG), Flare (FL) and Active Region (AR) to experiment. Every event has a detailed event boundary outline and timestamp which we used to generate the descriptor signatures. The summary of the solar event instances dataset is described in Table II.

TABLE II. Dataset structure

Wavelength	Event Types	Number of Event Instances
94	Sigmoid	208
94	Flare	20
131	Sigmoid	773
131	Flare	304
171	Active Region	1,167
171	Flare	200

Note that we selected each event and wavelength based on what the current SDO FFT modules use to identify the particular events as indicated in [1]. We selected these four different types of events since they are reported on AIA data and the FFT modules provide boudary outlines for them. This allows us to build a training set of properly identified regions of interest on the AIA images. A sample flare (FL) is shown on Figure 2.

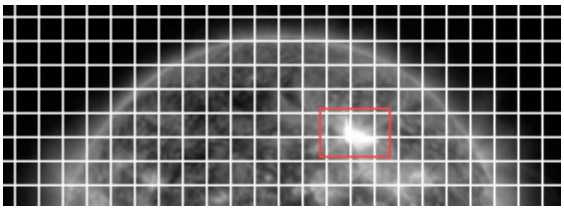

Fig. 2. AIA image section with the image parameter extraction grid overlaid and a flare (FL) event boundary box highlighted in light red

IV. Methodology

In this section, we discuss the process of generating the descriptor signatures using a histogram-based approach. We then introduce the six distance measures we will use to evaluate the performance of our signatures alongside our three test scenarios. Finally, we talk about the classification algorithms used for our independent evaluation.

A. Descriptor Signature Calculation

This is the most important and critial step of our work, since our region-based retrieval is built on the descriptor signatures. Our signature generation utilizes the pre-existing solar event labels which were reported by FFT modules and matched to the images in dataset [14]. Our descriptor signature is represented by a histogram-like structure with ten bins, one for each extracted image parameter and it is based on the average value of the image cells within the event's boundary. The procedure of signature calculation is described in Algorithm 1. This procedure matches each event to its corresponding cells, normalizes the values of each parameter on each selected cell with respect to the whole dataset, and finally averages the values of all cells of the same parameter, thus creating a histogram bin with each averaged value.

Algorithm 1 *Steps for calculating descriptor signatures*

1: Calculate the maximum value $Max(P_i)$ and minimum value $Min(P_i)$ of each of the 10 parameters, for all cells in the dataset. Where P_i is the i-th image parameter value.
2: Match the boundary outline of each event to the corresponding image cells. For each cell, find the parameter values.
3: Min-Max normalize each parameter value using: $P_i = \frac{P_i - Min(P_i)}{Max(P_i) - Min(P_i)}$
4: Take the average of each parameter and use it as a bin, in a histogram representing a given event.

After each signature is calculated, we will now have one ten-dimensional descriptor signature for each reported event. This considerably reduces the dimensionality of our previous approaches, were we had over 40,000 dimensions when comparing whole images as described in [2]. Region based descriptors will also allow this approach to only focus on important regions of each image. The selected regions are usually less than 10% of the entire image as shown in Figure 2.

B. Distance Measures for Retrieval Evaluation

To compare and match similar region-based solar events, we are going to pair-wise compare the different descriptor signatures. The results, when sorted, will generate our nearest neighbors lists for each event. Our experiments use several different distance measures to produce different nearest neighbor lists for performance comparison. The distance measures used to compare signatures are an important underlying factor that affects the querying results based on the peculiarities of each distance metric. For example, the cosine distance measures similarity based on the cosine of the angle between two vectors, while the euclidean distance does it by calculating the pair-wise distance between two elements of respective bins. Each of the measures selected are widely used in other content-based image analyses [5], [16], [17]. The formulas for these six measures are described in Table III.

TABLE III. DISTANCE MEASURES.

Distance Name	Formula
Euclidean Distance	$D1 = \sqrt{(x_s - x_t)(x_s - x_t)'}$
Cosine Distance	$D2 = 1 - \frac{x_s x_t'}{\sqrt{(x_s x_s')(x_t x_t')}}$
Jensen Shannon Divergence	$D3 = \sum_{j=1}^{n} A_j log \frac{2A_j}{A_j+B_j} + B_j log \frac{2B_j}{B_j+A_j}$
χ^2 Distance	$D4 = \sum_{j=1}^{n} \frac{A_j - B_j}{A_j + B_j}$
Kullback Leibler Divergence	$D5 = \sum_{j=1}^{n} A_j log \frac{A_j}{B_j}$

In euclidean distance x_s and x_t are the values of histogram bins. In cosine distance, the cosine of the angle between histogram bins x_s and x_t is calculated. In the rest of measures, A and B are a pair of histograms.

C. Test Scenarios

In our one month dataset, there are three different wavelengths. Images of each wavelength have two types of events, as described in Table II. Based on this fact, signature comparisons can be conducted in three scenarios, described in Table IV.

TABLE IV. TEST SCENARIOS FOR EVALUATION

Label	Scenario Description
S1	1-to-all comparison within the same type of events / wavelength combination
S2	1-to-all comparison within the same type of events
S3	1-to-all comparison of all signatures against each other (all events and wavelengths)

A more detailed explanation of what are we expecting to find with each different scenario is outline in Section V's subsections A, B, and C. Since each scenario is quite unique, we decided to present this information alongside the results of our experiments to enhance the reader's experience and understanding. At the end of each test scenario we will have an N-by-N similarity matrix as a result of the signature comparison among the N events. Each row (or column) of the matrix corresponds to an observation, with its distance to every other event instance. By sorting the values in each row by ascending order, we get the sorted nearest neighbors list for each event (observation). This list is the final retrieval result for each event.

D. Retrieval Accuracy Calculations

In this section we present our algorithm to calculate each event's retrieval accuracy for the experiments of scenarios S2 and S3. After we calculate the accuracy of all events we then average the results by event and/or wavelength to present them to the reader.

Algorithm 2 *Retrieval precision calculation*

1: Let E_i be the number of the ith event type ocurrences.
2: Calculate the top E_i nearest-neighbors of each event.
3: Determine how many of them are of the ith event type, call this true positives (TP).
4: Divide TP over E_i and multiply by 100. This results in the final accuracy percentage for that particular event.

E. Independent Evaluation using Classifiers

We use the WEKA data mining package to perform descriptor signature classification in order to verify our retrieval results. We will use four of the most popular classification algorithms: Naive Bayes, C4.5, Random Forest, and Support Vector Machines (SVM). These selected classifiers have been shown to produce solid results for solar data in the past [3], [4], [5]. We selected Naive Bayes and SVM (linear kernel) as our linear classifiers. To provide a broader scope, we also selected two decision tree-based classifiers: C4.5 and Random Forest. Both tree-based classifiers are fast to train. The SVM classifier, while expensive to train, is one of the most used ones in the literature thanks to its usually superior performance. All classification experiments have been run five times using 10-fold cross validation and we averaged all results.

V. EXPERIMENTAL EVALUATION RESULTS DISCUSSION

In this section, we will discuss the most interesting and relevant experimental results from each scenario. All the experiments were conducted using Matlab ©R2012a, and the classification experiments were performed with WEKA 3.4. Our source code and data can be found here **http://bit.ly/18zGLCr**, for anybody wanting to reproduce our results in the future.

A. Scenario S1 Results

In the most basic test, we calculate the distances between each type of event in the particular wavelengths chosen by the specialized FFT modules as the best to identify and label that event. In this experiment we expect to identify which wavelength / distance measure combinations are ideal to retrieve certain events, with our signature-based approach, and how this matches with what the FFT modules have determined over the past. One aspect of this scenario is observing how we can improve the separation of one event / wavelength combination by using different distance measures. We expect to see very different dissimilarity matrix plots for different combinations of events/wavelengths/distance measures. Our dissimilarity matrix plots are designed to look like heat maps where blue is more similar and red is the least similar. On the *X* and *Y* axis we have each event arranged by their time-stamp from earliest to latest. The diagonal of this plot should always be blue since the distance from each event to itself is zero. Each individual plot will showcase one particular type of event / wavelength at a time in order to visually inspect the how the distance measures enhance / mask their differences.

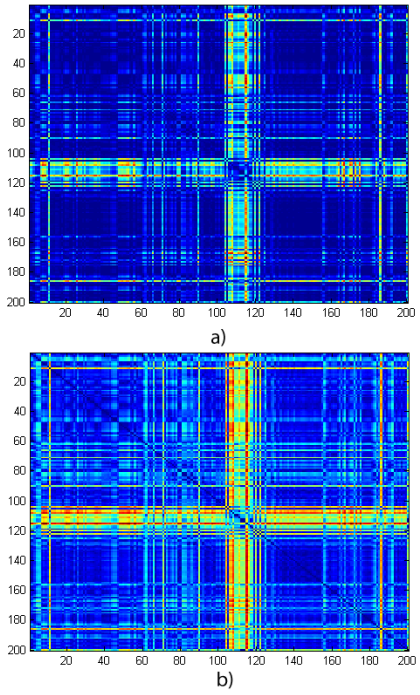

Fig. 3. Dissimilarity plots for S1 wavelength 171 - flare (FL) event with: a) Euclidean distance, b) Cosine distance

Figure 3 shows some of the benefits of using different distance metrics. On Figure 3(a) we used the Euclidean distance and we found that the FL event was not very easy to separate (with the clear exception of events 100 to 120. On the right (Figure 3(b)) we used the Cosine distance and things seem to separate better (notice the difference in colors), this indicates that the distances between signatures have changed and the results are easier to separate.

In order to contrast the retrieval differences between wavelengths, in Figure 4 we have one wavelength performing a lot better when it comes to identifying flares (171 Figure4(b)) since the colors clearly indicate dissimilar results in the ranges 100 to 140, in contrast with the results on Figure4(a), were there are almost no clear sections showing separations (light colors). Interestingly enough, this wavelength (171) is actually preferred to identify active regions (AR) more than flares (FL) as found in [18].

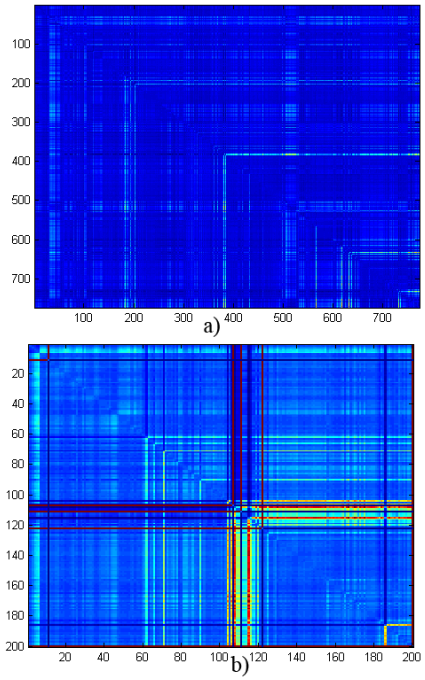

Fig. 4. Dissimilarity plots for S1 wavelengths (a) 131 and (b) 171 for the flare (FL) event. Both plots represent Kullback Leibler Divergence measure

While we have 28 more plots, these ones presented here (Figures 3 and 4) are some of the most interesting and useful results. To see all the additional results, please check **http://bit.ly/18zGLCr**.

B. Scenario S2 Results

Moving a bit away from the visual analysis provided by the dissimilarity matrix plots from S1 (Figures 3 and 4), we now provide our commentary based on the retrieval precision calculated by Algorithm 2, where we are using our histogram-based distance measures. In this section we will show retrieval precision results that further validate the findings by the FFT modules and strengthen our previous results.

TABLE V. SIGMOID (SG) RETRIEVAL PRECISION

Wavelength	Distance	Precision
94	KLD	91.48%
131	KLD	42.24%

As we can see in Table V, we are retrieving 91.48% of our sigmoid (SG) events correctly on the proper wavelength that the sigmoid (SG) detection module is using. This validates their science results as well as ours when using our descriptor signature region-based querying.

The results in Table VI continue to match with the FFT science modules. The best retrieval precision was obtained with the 131 wavelength, which is the preferred wavelength of the

flare (FL) finding module. While this result shows that we are not always able to find new wavelength / event combinations for detection (using our signature-based approach), but we can still rely on our previous expert knowledge (FFT modules) to make the proper wavelength selection for detection on our retrieval system. This also confirms our fears that when the events are combined on one wavelength we might not be able to find them as effectively. A clear example of this is Table VII .

TABLE VI. FLARE (FL) RETRIEVAL PRECISION

Wavelength	Distance	Precision
94	KLD	14.50%
131	KLD	75.50%

In order to show how we can combine both S1 and S2 to gain valuable insights, we provide the following analysis for our wavelength 171. In the next table we present the retrieval precision results (with the Euclidean distance) for the flare (FL) and the active region (AR) events generated in S1 in Table VII.

TABLE VII. FLARE (FL) AND ACTIVE REGION (AR) RETRIEVAL PRECISION FOR THE EUCLIDEAN DISTANCE

Wavelength	Event	Precision
171	Flare (FL)	24.17%
171	Active Region (AR)	84.92%

These last results help us confirm that when the events are combined into the same wavelength we match the expert knowledge (FFT module) of where to find each event by retrieving them with high precision. This also shows that we have a clear advantage findings flares (FL) in wavelength 131 (Table VI) than in wavelength 171 (Table VII), as also seen visually, in 4(a) they all look the same (dark blue plot), making them easier to separate from other events within the same wavelength.

C. Scenario S3 Results

By combining all different wavelengths and labels we can now provide general retrieval results for all events using different distance measures.

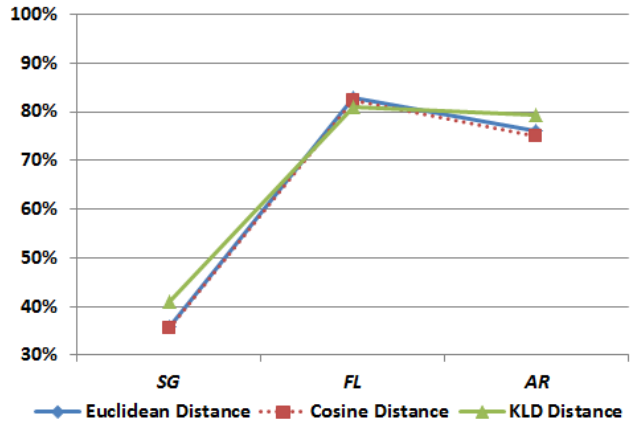

Fig. 5. Retrieval Precision for All Events Using Euclidean, Cosine, and KLD Distance Measures

As we can see from Figure 5, we do a fairly good job in identifying flares (FL) and active regions (AR) in general, but we are still very ineffective at finding sigmoids (SG) in general context, we theorize this is due to the fact we are using bounding boxes that add more noise to the sigmoid events than the flare events. The best performing measure on average was the KLD histogram-based measure, and even that measure only correctly identifies 40% of the sigmoids (SG). In a general context (S3), this result (for sigmoids (SG)) is quite disappointing. However, we did achieve good results (91%) using KLD (as seen in Table V with the wavelength 94, in the S2 experiments. The combination of results from our S1, S2, and S3 experiments show that we can define a set of rules in a more flexible and all-purpose system when compared to the other FFT science modules. Note that these retrieval results only differentiate events between each other, we are not considering Empty sun or multi-labeled regions in this work.

D. Independent Evaluation Results based on Classification

In order to have a non-image retrieval verification of our results and descriptor signature approach in general, we decided to use four different classification algorithms. We used the same data as in Scenario S3, i.e. all wavelengths grouped by event labels. By analyzing these results we expect to find similar findings to our previous sections, thus validating our descriptor signature approach. If we can classify the descriptor data somewhat correctly, then we should be on track to be able to use these signatures to classify our solar events properly. This will let us independently reproduce classification performed by other FFT modules, in an all-purpose module.

Our general results are presented in Figure 6. We present these results ordered by our classification algorithms.

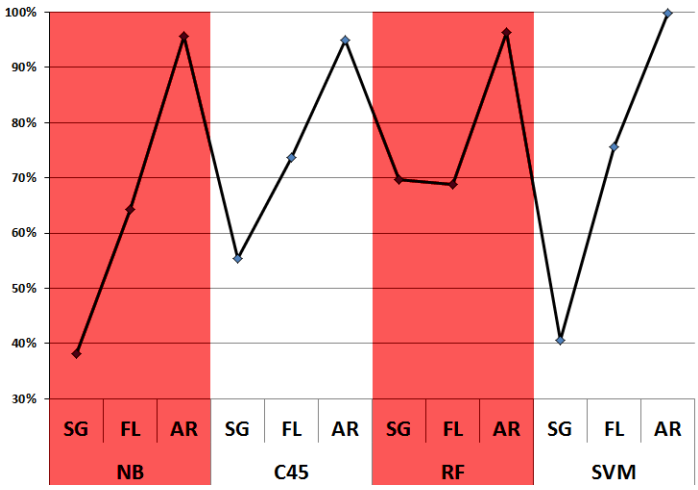

Fig. 6. Classification Accuracy of Solar Events

As expected, in Figure 6 we show the classification results for our three classifiers next to each other for the all wavelengths experiment (S3). We use this figure to show, on average, the very poor performance for sigmoids (SG). This leads us to believe that only with the KLD distance we are able to get good performance out of our descriptor signatures while only working on a single wavelength. The most interesting

behavior in this figure is that both tree-based classifiers (C 4.5 and Random Forest (RF) see a very considerable improvement for sigmoids (SG), improving from 40% accuracy to nearly 70% accuracy (for Random Forest (RF)), something acceptable and quite interesting. We believe this is mostly due to the nature of the classifier. We also notice that active region (AR) almost nears 99% accuracy for all classifiers, making it the easiest event to classify.

Note that we have left our classes unbalanced in order to fully mimic the reality of our global SDO data, where certain events appear more than others. While active region (AR) classification may be more accurate because they are the most common event, they are only slightly more numerous than sigmoids (SG), which performed the worst. Flares (FL) perform fairly consistently, and they come in third for numbers of events reported. In the future, we intend to gather more labeled data in order to create a balanced dataset with enough data samples to verify our findings in this work. The combination of events across wavelengths is intended for possible identification of solar events on different wavelengths (other than the ones used in the FFT modules) with our image parameters, an approach that is different than what the FFT modules are doing and the premise of our trainable module.

VI. Conclusions

Having found plenty of interesting results, some contradictory at a glance and others that very strongly match the FFT science modules findings with our current approach, using descriptor signatures is a very promising avenue of research for our purposes and work with solar physics data. We have shown that by following the traditional wavelength / event pairings we get solid retrieval accuracy (over 75% for all events), and while we are still not very accurate on detecting sigmoids (SG) (with our approach), we are able to rely on the expert knowledge of the Sigmoid Sniffer Module for now [1]. We have also shown that when combining events in different wavelengths we can see some interesting results, and while they did not really come to fruition with this particular dataset and distance measure combination, there is still the potential that other events and/or other measures will produce something more interesting. The most interesting findings, in our opinion, are the ones achieved when combining the wavelengths and by just considering the events by themselves. These results allowed us to analyze their interactions using traditional retrieval mechanisms and classification algorithms. With our approach the retrieval calculations can now be done in real time due to the compactness of descriptor signatures. Thus allowing the system to not rely on large pre-computed nearest-neighbor (NN) tables as it currently does.

VII. Future Work

As one of our immediate plans, we expect to gather a bigger dataset with emphasis on acquiring more training samples for the sigmoid (SG) class, and experimenting with balanced and boosted classes. We also expect to add more event types to determine how well we can retrieve them using our current methodology. Since our current methodology works effectively, we will provide these capabilities on our actual CBIR system in the coming months to have researchers test them and explore their power.

Since we have shown during our work while building our CBIR system that our image parameters produce very good results when used to retrieve medical images [19], we plan to look into expanding these region-based querying capabilities to other image domains, including images from the Galaxy Zoo project (http://www.galaxyzoo.org) which seems to have similar characteristics to our solar events. We encourage astrophysicists to present us their datasets so we may apply some of the methods used in this work, or our previous works, to help them with their large-scale retrieval and classification issues.

VIII. Acknowledgments

This work was supported by two National Aeronautics and Space Administration (NASA) grant awards, 1) No. NNX09AB03G and 2) No. NNX11AM13A.

References

[1] P.C.H. Martens, G.D.R. Attrill, A.R. Davey, Engell, A, *et. al*. Computer vision for the solar dynamics observatory (sdo). *Solar Physics*, 275:79–113, 2012.

[2] J.M. Banda, R.A. Angryk, and P.C.H. Martens. On dimensionality reduction for indexing and retrieval of large-scale solar image data. *Solar Physics*, 283:113–141, 2013.

[3] J.M. Banda and R.A. Angryk. An experimental evaluation of popular image parameters for monochromatic solar image categorization. In *23rd Inter. FLAIRS Conf.*, pages 380–385, 2010.

[4] J.M. Banda and R.A. Angryk. On the effectiveness of fuzzy clustering as a data discretization technique for large-scale classification of solar images. In *Proceedings of the 18th IEEE International Conference on Fuzzy System*, pages 2019–2024, 2009.

[5] J.M. Banda and R.A. Angryk. Selection of image parameters as the first step towards creating a cbir system for the solar dynamics observatory. In *Digital Image Computing: Techniques and Applications (DICTA)*, pages 528–534, 2010.

[6] J.M. Banda, R.A. Angryk, and P.C.H. Martens. Steps toward a large-scale solar image data analysis to differentiate solar phenomena. *Solar Physics*, pages 1–28, 2013.

[7] R. Chakravarti and X. Meng. A study of color histogram based image retrieval. In *Information Technology: New Generations. Sixth International Conference*, pages 1323–1328, 2009.

[8] A. Pentland, R Picard, and S Sclaroff. Photobook: Content-based manipulation of image databases. *International Journal of Computer Vision*, 18(3):233–254, 1996.

[9] M. Flickner, H. Sawhney, *et. al*. Query by image and video content: The qbic system. *Computer*, 28(9):23–32, 1995.

[10] P.M Kelly, T.M Cannon, and D.D Hush. Query by image example: the comparison algorithm for navigating digital image databases approach. In *Storage and Retrieval for Image and Video Databases*, pages 238–248, 1995.

[11] V.E. Ogle and M Stonebraker. Chabot: Retrieval from a relational database of images. *Computer*, pages 40–48, 1995.

[12] C. Carson, M. Thomas, *et al*. Blobworld: A system for region-based image indexing and retrieval. In *Visual Information and Information Systems*, pages 509–517, 1999.

[13] J. Lemen, A. Title, D. Akin, *et al* The atmospheric imaging assembly (aia) on the solar dynamics observatory (sdo). *Solar Physics*, 275(1-2):17–40, 2012.

[14] M.A. Schuh, R.A. Angryk, K.G. Pillai, J.M. Banda, and P.C.H. Martens. A large-scale solar image dataset with labeled event regions. In *20th IEEE Int. Conf. on Image Processing (ICIP)*, pages 4349–4353, 2013.

[15] H. Tamura, S. Mori, and T. Yamawaki. Textural features corresponding to visual perception. *Systems, Man and Cybernetics, IEEE Transactions on*, 8(6):460–473, 1978.

[16] W Cao, Z. Shi, and J. Feng. Traffic image classification method based on fractal dimension. In *Cognitive Informatics, 5th IEEE International Conference*, pages 903–907, 2006.

[17] V. Devendran, T. Hemalatha, and W. Amitabh. Svm based hybrid moment features for natural scene categorization. In *International Conference on Computational Science and Engineering*, pages 356–361, 2009.

[18] V. Delouille, B. Mampaey, C. Verbeeck, and R. de Visscher. The spoca-suite: a software for extraction and tracking of active regions and coronal holes on euv images. In *Arxive-prints, 1208.1483*, 2012.

[19] J.M. Banda, R.A. Angryk, and P.C.H. Martens. On the surprisingly accurate transfer of image parameters between medical and solar images. In *18th IEEE Int. Conf. on Image Processing*, pages 3669–3672, 2011.