On Using SIFT Descriptors for Image Parameter Evaluation

Patrick M. McInerney, Juan M. Banda
Department of Computer Science
Montana State University
Bozeman, Montana 59717
patrick.mcinerney, juan.banda@cs.montana.edu

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

***Abstract*—In this work we present a composite method for image parameter evaluation using Scale-Invariant Feature Transform (SIFT) descriptors and bag of words representation applied to pre-selected image parameters, with potential applications to solar data and other domains. As one of the main challenges in computer vision, image parameter evaluation has been approached from supervised and unsupervised perspectives. Taking advantage of the SIFT scale and rotation invariant properties; we propose a combined method that will aid the image parameter selection process when applying SIFT and bag of words on top of pre-selected parameters. We provide a comparison against traditional methods and across several different datasets to validate our method.**

Keywords: Image retrieval, image descriptors, classification

I. Introduction

As solar physics enters the era of big data, modern machine learning and data mining techniques becomes increasingly important in managing and learning from the massive and growing stream of solar information. Automated detection and analysis of solar phenomena is necessarily becoming the norm, and a great deal of continued work is needed to improve and refine the first generation of solar information processing tools [1].

The desire to overcome the limitations of heavily specialized existing systems has produced the first general purpose solar Content-Based Image Retrieval (CBIR) system, developed and hosted by the data mining lab of Montana State University[1]. This system breaks images of the Sun produced by NASA's Solar Dynamics Observatory (SDO) down into basic visual attributes, and these attributes can then be used to compare and return images similar to some image of interest. These attributes were carefully selected as appropriate for solar images, and some attribute evaluation analysis has already been performed with them [2]. However, the system still stands to benefit greatly from new or improved methods and techniques better suited to its data.

The quality of data attributes is a critical component of any machine learning or retrieval system. If the system learns from attributes that are noisy, irrelevant, or redundant, it will have a much more difficult time finding generalizable methods of describing its data. Even with relevant attributes, there may be a desire to retain and use only those most useful to the learning task in order to reduce the storage space or computational power required for operation. Thus, attribute selection techniques must be selected and used to refine the attribute set and improve the operation of the system.

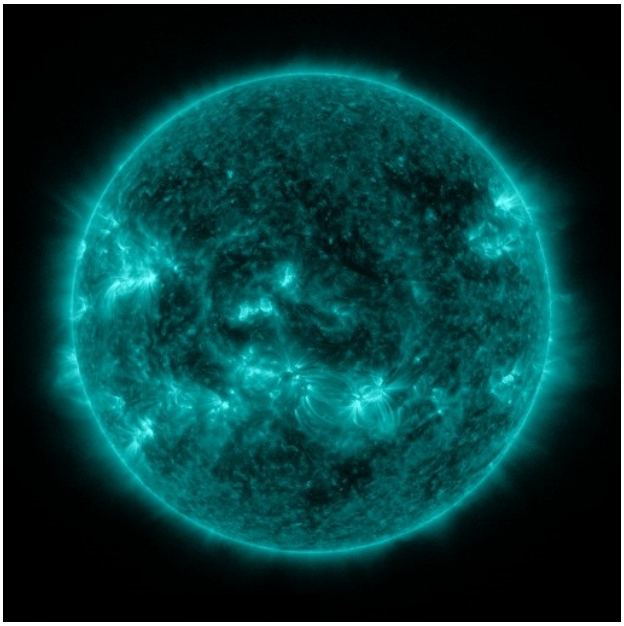

Fig. 1. A sample solar image taken by the SDO mission. Image taken from http://sdo.gsfc.nasa.gov/

A wide range of parameter evaluation methods exist to facilitate attribute selection. However, benchmark comparisons of many attribute selection methods [3, 4] have demonstrated that the success of attribute selection techniques is highly dependent upon how they interact with the learning algorithm used and the characteristics of the data in any given application.

Solar image retrieval poses its own unique difficulties for parameter selection due to the distinctive structure of the data. The inherent spatial relation of visual features in an image often leads to a desire for spatially aware image parameters. The retrieval method we study in this paper, based on the one used by the SDO CBIR system, achieves this spatial representation by extracting its image parameters on

[1]Available at http://cbsir.cs.montana.edu/sdocbir/

a grid-based segmentation of the image. This produces many spatially related samples of a single parameter for each image. Unfortunately, the most important classical attribute evaluation measures, such as information gain or the χ^2 statistic, are poorly equipped to capture spatial relationships in their input parameters, and much of this valuable information is lost in their use.

For this reason we present a new method for attribute selection in such a system, based on SIFT descriptor extraction from the grid of parameter values. Although SIFT is more typically used as a base level image attribute, we demonstrate here that it can also achieve strong results in parameter selection on images in solar and other domains, indicating promise as a useful way to select image attributes in spatial contexts such as the SDO CBIR grid-based segmentation representation.

The remainder of this paper is structured as follows. Section II provides some background material in solar computer vision. Section III defines and characterizes the image parameters used to represent our images. Section IV introduces the image datasets used in our experiments. Section V develops our SIFT-based parameter selection technique. It also describes the classical attribute selection methods against which our technique is compared. Section VI details the methodology of our experiments. Section VII presents and analyzes the results of these experiments. Section VIII concludes, and Section IX offers avenues for future work.

II. Background

With the advent of the SDO mission, solar physics has gained a need for big data processing techniques. Launched on February 11, 2010 as the first mission of NASA's Living With a Star program, the SDO captures full-disk imagery of the sun in unprecedented detail. Combining high resolution instrumentation with a rapid image capture cadence, the mission will produce more data than all previous solar missions combined [1]. Consequently, automated detection and characterization of solar events must necessarily replace the manual analysis that has been typical in the research of solar phenomena.

To this end, an international consortium of independent groups, titled the SDO Feature Finding Team (FFT), was assembled to produce a set of computer vision modules capable of recognizing important solar phenomena in SDO data [1]. Fifteen of these sixteen modules are designed to detect and, in some cases, categorize one or two classes of well-studied solar phenomena, such as coronal holes or solar flares. These modules rely extensively on foreknowledge of the visual characteristics of the targeted event type for successful operation. Consequently, identifying new event types or notable varieties of existing event types would require the substantial expense of either creating new modules or heavily modifying existing ones.

The sixteenth module, built by our interdisciplinary research group at Montana State University, is a trainable module, designed to work in conjunction with a general CBIR system for solar images [5]. The benefits of a general system include the ability to compare phenomena of any class to find those which are visually similar, as well as to provide a way to search for novel or uncommon event types or solar regions, which will not have a dedicated detection module.

Previous work on this General CBIR system evaluated the performance of a variety of image parameters on solar images [2, 6]. To accomodate the need of continuous calculation of these parameters in the SDO data stream, the ten most appropriate parameters were selected based on a combination of classification accuracy and computational efficiency. These parameters were able to successfully differentiate localized images of specific events [7], as well as identify filaments in the full disk of the Sun [8]. It was also discovered that image parameters found to work well for medical x-ray images also worked well for solar images [9], which suggested significant visual similarity in these images, and therefore potential cross-application of image retrieval techniques between the two domains.

III. Image Features

The parameter evaluation experiments of this paper use the ten texture-based image features of the SDO CBIR system [5]. The utility of specific image features for any image processing or retrieval task is highly domain dependent, and these parameters were chosen from the vast field of options based on the specific characteristics of solar imagery. In particular, the monochromatic nature of the images and patterned variation in the intensity of solar radiation led to a focus on texture-based parameters. These same monochrome and patterned variation properties apply to medical imagery, and early exploration found that parameters commonly used for medical image classification and retrieval also worked quite well in retrieval of solar images [9]. The parameter selection process was consequently able to draw on the well developed CBIR work that has been done in the medical domain [10]. The parameters chosen and used by the system are mean, standard deviation, skewness, kurtosis, entropy, relative smoothness, uniformity, Tamura's contrast and directionality features, and fractal dimension, all calculated from pixel intensity [2, 6]. The formulas for these parameters are given in Table I.

TABLE I
Texture Parameter Formulas

Number	Parameter	Formula
P1	Mean [11]	$m = \frac{1}{L}\sum_{i=0}^{L-1} z_i$
P2	Standard Deviation [11]	$\sigma = \sqrt{\frac{1}{L-1}\sum_{i=0}^{L-1}(z_i - m)}$
P3	Skewness [11]	$\mu_3 = \sum_{i=0}^{L-1}(z_i - m)^3 p(z_i)$
P4	Kurtosis [11]	$\mu_4 = \sum_{i=0}^{L-1}(z_i - m)^4 p(z_i)$
P5	Entropy [11]	$-\sum_{i=0}^{L-1} p(z_i)\log_2 p(z_i)$
P6	Uniformity [11]	$-\sum_{i=0}^{L-1} p^2(z_i)$
P7	Relative Smoothness [11]	$1 - \frac{1}{1+\sigma^2(z)}$
P8	Tamura Contrast [12]	$\frac{\sigma^2}{(\mu_4)^{0.25}}$
P9	Tamura Directionality [12]	See description in Section III
P10	Fractal dimension [13]	$D_0 = -\lim_{\epsilon\to 0}\frac{\log N(\epsilon)}{\log(\epsilon)}$

L = Number of pixels in image segment
z_i = Intensity of pixel i
$p(z_i)$ = Histogram value of intensity level z_i
$N(\epsilon)$ = number of size ϵ boxes needed to cover the area

Mean, standard deviation, skewness and kurtosis are standard statistical measures. Entropy measures how complex a binary signal is required to encode a piece of data. It will increase if there is a high variety of intensities in an image or image segment, and will approach zero as the intensities become less varied. Relative smoothness is a measure of contrast. Images consisting entirely of one pixel intensity will have a smoothness of zero, while images with many very different pixel intensities will have a high smoothness value. The uniformity feature ranges from one to zero, and is larger when the distribution of pixel intensities is concentrated in only a few values. Tamura contrast and directionality are two of six features, introduced in [12], attempting to capture human visual perception numerically. Tamura contrast tries to represent the dynamic range polarization of the grey levels of the image. Tamura directionality measures the degree to which changes in intensity of the image or image segment occur in the same direction. It does this by first constructing a histogram over the gradient values of the image, binned by direction, then calculating the the second moment of the histogram around the maximum histogram value, producing a single numerical representation. Fractal dimension is a measure designed to give an indication of how fine-grained a fractal's structure is. It is also a useful way to measure the complexity or visual roughness of any image. We estimate the fractal dimension using the standard box counting method.

IV. Datasets

In addition to our interest in the usefulness of our method for solar CBIR, we want to explore and validate the utility of our method outside of the solar domain. Therefore, our experiments are performed using a collection of eight image datasets, representing a broad range of image retrieval environments, including solar, medical, indoor environment, and general object recognition domains. Each image collection we use is sampled to produce a set of evenly balanced classes for our experimental dataset. Each image belongs to exactly one class, and each class label is applied only to entire images, e.g. a *bicycle* class image would not have the part of the image containing the bicycle delimited in any way. Datasets with color images are converted to greyscale. An overview of these datasets is presented in Table II. Sample images from each dataset are given in Figure 2. For further detail on the construction of the ImageCLEFmed, PASCAL, INDECS, and TRACE datasets, please refer to [9].

TABLE II
Dataset Properties

Dataset	Domain	#Classes	Images per class
TRACE [14, 15]	Partial Sun	8	200
AIA [16]	Full Sun	6	298
MSRCORID [17]	Natural Scene	8	200
PASCAL 2006 [18]	Natural Scene	8	200
PASCAL 2008 [18]	Natural Scene	8	200
INDECS [19]	Indoor	8	200
CLEFMED 2005 [20]	Medical	8	200
CLEFMED 2007 [20]	Medical	8	200

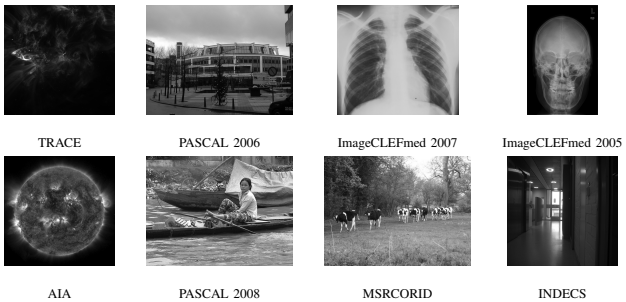

Fig. 2. Sample images of our experimental datasets

A. TRACE

The TRACE dataset we use [14, 15] is a selection of labeled images from the Transition Region and Coronal Explorer (TRACE) satellite, which was in operation from 1998-2010, taking pictures of the solar atmosphere. In contrast to the full disk images of newer SDO dataset we also use, this collection contains only partial disk images of the Sun. These partial sun images each focus on an area containing a single solar phenomenon. All images are 1024 × 1024 pixels in size. Most of the images come from different event occurrences. However, in some cases where events occur infrequently, but for a prolonged period of time, multiple images of a single event at different periods of its lifetime may be included.

B. AIA

Our second solar dataset is taken from a collection of solar data from the SDO [16]. This collection contains a set of solar events in six categories reported by the customized SDO FFT computer vision tools between January 1 and June 30 2012. It also contains the SDO CBIR parameters extracted from a set of 4096 × 4096 pixel images of the Sun taken by the SDO's Atmospheric Imaging Assembly (AIA) instrument during that time, chosen to best capture the reported events. Our experimental dataset built from this collection takes the events reported in January, and randomly undersamples the more frequent events such that each event has the same number of reports selected. One representative image is chosen for each selected event report.

C. MSRCORID

The Microsoft Research Cambridge Object Recognition Image Database (MSRCORID) is a publicly available image collection intended for object recognition research [17]. The images are broken down into 18 categories, covering a varied range of everyday objects, such as cars, flowers, signs, or buildings. Each image is either 480 × 640 pixels or 640 × 480 pixels in dimension. Our experimental dataset draws images from the 8 largest MSRCORID categories: animals, bicycles, cars, chimneys, clouds, scenes, trees, and windows.

D. PASCAL

The PASCAL Visual Object Classes (VOC) challenges, running from 2005 to 2012, provided images of realistic scenes and tasked machine learning systems with identifying the

presence of everyday objects such as people, cats, airplanes, or television sets in a collection of images. We make use of the 2006 and 2008 datasets from this series [18]. The 2006 collection contains 10 object classes for its images, and the 2008 collection has these 10 classes plus 10 additional ones, for a total of 20. The 2008 collection also uses a new, larger set of images. The resolution of these images is variable, generally 300-500 pixels in each dimension.

E. INDECS

The INDECS (INDoor Environment under Changing conditionS) collection consists of pictures taken in five rooms of a typical office building from various viewpoints [19]. The classes of the collection consist of the combination of the room and the lighting condition under which the picture was taken, such as Corridor-Night, or Printer area-Cloudy. The images are taken under normal use conditions, including the presence of people. All images were originally 1024 x 768 pixels, but were resized to 1024 × 1024 pixels for our experiments to better facilitate our segmentation procedure.

F. ImageCLEFmed

Conference and Labs of the Evaluation Forum [2] (CLEF) is a conference and body promoting multilingual and multimodal information retrieval research. We use two datasets drawn from the medical component of their image track, ImageCLEFmed [20]. Sampled from the 2005 and 2007 versions of their image collection, the images in these datasets are radiographs of the human body, manually annotated for machine learning use. In 2005, the collection drew images from four different medical sources. In 2007, two additional sources were added and the annotations were modified. Note that although the 2005 dataset is a subset of the 2007 dataset, our samplings are disjoint of each other. The resolution of these images is variable, but in all cases at least one dimension is 512 pixels.

V. Attribute selection techniques

A. SIFT bag of words representation

In our parameter selection technique, we utilize the Scale-Invariant Feature Transform (SIFT) descriptor to capture some of the spatial qualities of our parameter grid. SIFT features are a local image description tool, designed to be robust against changes in scale, rotation, and translation of the described area [21]. To extract a SIFT description of an image, key points are found for the image that are stable and invariant to scale and orientation changes. Next a 128-dimensional SIFT descriptor is calculated for each key point found. This descriptor captures local gradient values around the key point at the perceived scale of the feature.

Comparing arbitrary images using the raw SIFT descriptor representations can be difficult and computationally intensive for images with many descriptors. Drawing on previous work in SIFT-based methods [22], our image characterization creates a bag of words representation on top of the SIFT descriptors to alleviate this problem. First adapted from text recognition techniques, bag of words representations count the instances of a feature or word present in a text, ignoring any ordering that may be present. In this case, the 'text' is an image and the 'words' are formed by taking the output centers after clustering the SIFT descriptors. Each descriptor found in the image is mapped to the nearest center to produce a word count for each visual word. These counts form the vector used to represent the image.

B. Classical attribute selection techniques

We test our method against three commonly used measures for attribute selection: the χ^2 measure, the information gain measure, and the gain ratio measure.

The χ^2 test [23] is a statistical test used to determine the statistical independence of two events or variables. In the context of attribute selection, the test is used to determine the independence between the value of a parameter and the value of the class. If the test says that the parameter and the class are highly unlikely to be independent, then the parameter is considered a strong candidate for selection.

The information gain measure, also called the Kullback-Leibler divergence [23], captures how much the entropy of the class is reduced given knowledge of the specified attribute. Attributes with high information gain are better able to capture all the complexity of a class distribution than those with low information gain, and therefore are likely good candidates for attribute selection.

The gain ratio measure [24] calculates the ratio of an attribute's information gain and its intrinsic value, a measure of the complexity of the attribute. information gain tends to prefer complex attributes with many values, which can result in overfitting. Gain ratio compensates for this by essentially penalizing attributes that can adopt many different values, and consequently can be a better method for choosing the most useful attributes.

VI. Experimental Setup

A. Parameter Extraction

To produce the texture parameter representations of our images, we follow the methods developed and used in the existing SDO CBIR system. First, we split each image as evenly as possible into a 32 × 32 grid of image segments. We then calculate each of the ten parameters described in Section III on each segment, and store these extracted parameters as a set of ten 1024 dimensional single-parameter vectors for each image. In the case of the SDO dataset, whose images are much larger than the others, a 64 × 64 grid is used instead.

Once these individual parameter representations have been constructed, their classification performance is calculated using the WEKA machine learning software package [25]. Each set of single parameter data is used as input for training and testing a classifier to provide a baseline against which we measure the performance of our attribute selection methods. We test our image with two different classifiers: a Support

[2]formerly Cross-language Evaluation Forum

Vector Machine (SVM) and a Random Forest (RF) classifier. An SVM attempts to create an optimal hyperplane that separates two classes within the parameter space or within some mapping of the parameter space to a higher dimensional space. The optimality is judged based on the distance from the closest point of either class to the hyperplane. An RF classifier is an ensemble method that constructs a number of different decision trees, and then outputs the most common classification given among the output of those trees. The use of multiple classifiers allows us to obtain broader insight and a more complete analysis of the performance of our attribute selection methods. These classifiers are widely used and have strong, well established performance in many domains, making them natural choices for our experiments.

B. Attribute selection

Our SIFT attribute evaluation (SIFT AE) method computes a SIFT bag of words representation of the parameter data extracted on the segment grid. Treating each image's extracted data for a given parameter as a 32 × 32 pixel image (or in the case of the SDO dataset, 64 × 64), we extract SIFT descriptors as described in Section V. Next, we take those descriptors and perform k-means clustering on them. These k cluster centers then become the 'words' of the representation. Every descriptor found is mapped to the nearest word, adding to its count for that image. The total count of each word then becomes the vector representation of the image. We perform our experiments with k = 20, 50, and 100. SIFT feature extraction, and the construction of bag of words representations are performed using the tools of the VLFeat open source computer vision library [26].

To compute a parameter selection ranking with this method, we take the parameter SIFT bag of words representations for our given dataset and evaluate their classification success in the same manner as the basic parameters described above. We average the classification accuracies from the three different word counts, then order the parameters by decreasing average classification accuracy.

We compare our SIFT AE rankings against three standard attribute selection tools available within the WEKA framework: χ^2, information gain, and gain ratio. When using these methods, each image segment is treated as a separate sample from the dataset, and the method is applied to the set of segments. We present the ranking of each parameter as returned by the WEKA implementation of the methods.

To evaluate each of the four measures, we compare the ranking produced by the measure against the correct ranking of parameters by their actual classification performance. This ranking is considered correct because the best parameter for the purposes of parameter selection is the one that will produce the best classification results. This comparison is quantified by using the the inversion number (IN) measure [27]. The inversion number is the number of instances across all pairs of attributes where the pair's relative ordering in the measure-produced ranking is incorrect with respect to the classification-based ranking. A lower inversion number means the measure's

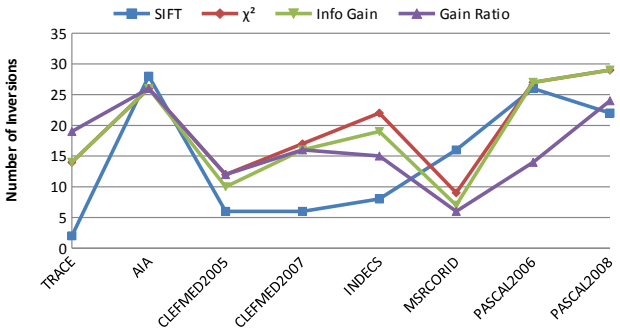

Fig. 3. Inversions plot for SVM classification

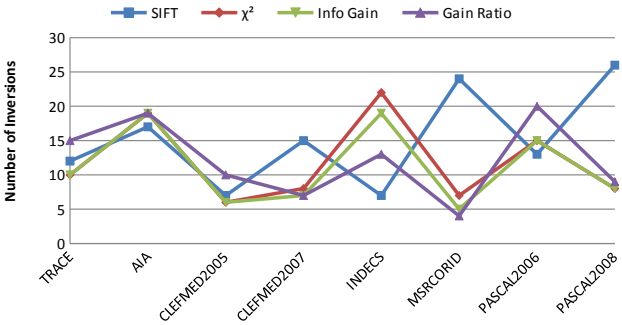

Fig. 4. Inversions plot for RF classification

ranking is closer to the classification ranking, and therefore is performing better. For ten parameters, the maximum possible inversion number is 55, which occurs when the measure ranks the parameters in the complete opposite order of their classification success. Note that the four methods we compare and our ranking evaluation all focus on single parameter selection. When selecting multiple parameters to use, the best individual parameters may not produce the best classification results of all possible combinations.

VII. Results and Analysis

Figures 3 and 4 show the inversion numbers for all of our evaluations. We next present a closer analysis of most interesting and illustrative results from our experiments. In each of the following tables, the first column lists the rank numbers, the second column lists the correct ranking, i.e. the list of parameters ordered by their accuracy when used as input to the classifier. The next four columns list the ranking produced by each attribute selection method. Above each ranking is the IN for that method.

Tables III and IV show the experimental results for the TRACE dataset. These results show quite clearly that the classical methods are not effective at selecting useful parameters for some datasets, and that our SIFT-based method can be much more effective.

Not only does our method return a ranking with an extremely low inversion number, much lower than the other methods, but the other methods rank some parameters very far from their correct position, which would be very detrimental to attribute selection. All three classical methods fail to detect mean (P1) as one of the best parameters for classification, and

TABLE III
RANKING COMPARISON WITH INVERSION NUMBERS FOR SVM CLASSIFICATION OF TRACE DATASET

Rank	Parameter	SIFT AE IN: 2	χ^2 IN: 14	Gain Ratio IN: 19	Info Gain IN: 14
1	P1	P1	P7	P10	P7
2	P5	P5	P6	P6	P6
3	P3	P3	P5	P7	P5
4	P6	P7	P3	P5	P3
5	P7	P4	P4	P3	P4
6	P4	P6	P10	P4	P10
7	P2	P2	P1	P8	P1
8	P10	P10	P8	P1	P8
9	P8	P8	P9	P9	P9
10	P9	P9	P2	P2	P2

TABLE IV
RANKING COMPARISON WITH INVERSION NUMBERS FOR RF CLASSIFICATION OF TRACE DATASET

Rank	Parameter	SIFT AE IN: 12	χ^2 IN: 10	Gain Ratio IN: 15	Info Gain IN: 10
1	P6	P1	P7	P10	P7
2	P7	P5	P6	P6	P6
3	P1	P4	P5	P7	P5
4	P3	P6	P3	P5	P3
5	P4	P3	P4	P3	P4
6	P5	P7	P10	P4	P10
7	P8	P2	P1	P8	P1
8	P2	P10	P8	P1	P8
9	P10	P8	P9	P9	P9
10	P9	P9	P2	P2	P2

the gain ratio technique identifies fractal dimension (P10), one of the worst performing parameters, as the best. Standard deviation (P2) is uniformly ranked last by all classical methods, but in fact performs much better, ranking fifth in the actual classification results.

TABLE V
RANKING COMPARISON WITH INVERSION NUMBERS FOR SVM CLASSIFICATION OF IMAGECLEFMED 2007 DATASET

Rank	Parameter	SIFT AE IN: 6	χ^2 IN: 17	Gain Ratio IN: 16	Info Gain IN: 16
1	P1	P1	P6	P6	P6
2	P5	P5	P7	P7	P7
3	P3	P3	P1	P1	P1
4	P4	P4	P2	P2	P2
5	P2	P7	P8	P5	P5
6	P6	P6	P5	P8	P8
7	P7	P9	P3	P3	P3
8	P8	P2	P4	P4	P4
9	P9	P10	P10	P10	P10
10	P10	P8	P9	P9	P9

The ImageCLEFmed 2007 dataset (see Table V) is another example of traditional methods not reflecting the best performing parameter. The best attributes, as reported by all three methods, are uniformity (P6) and relative smoothness (P7), but these parameters are of median accuracy in practice. Our method again identifies the best performing parameters (mean and entropy) correctly, and again achieves a significantly lower inversion score than any of the classical methods. The other medical dataset (ImageCLEFmed 2005) offered very similar results.

The similarly successful behavior of our method on the TRACE and medical datasets agrees with the knowledge that these images are very similar in character [9]. Intriguingly, we also found strong performance of our method on the INDECS dataset (see Tables VI and VII). Although this dataset is of very different character, it also had a much lower inversion number under SIFT AE than the others, behavior that was consistent under both the SVM and RF classifiers. This indicates that our method is not limited to one type of data, but is applicable to other datasets as well.

TABLE VI
RANKING COMPARISON WITH INVERSION NUMBERS FOR SVM CLASSIFICATION OF INDECS DATASET

Rank	Parameter	SIFT AE IN: 8	χ^2 IN: 22	Gain Ratio IN: 15	Info Gain IN: 19
1	P3	P1	P2	P5	P2
2	P4	P4	P5	P6	P5
3	P1	P7	P8	P7	P7
4	P5	P5	P7	P3	P6
5	P7	P3	P6	P2	P3
6	P6	P2	P3	P4	P8
7	P2	P6	P4	P8	P4
8	P9	P8	P1	P1	P1
9	P8	P9	P10	P10	P10
10	P10	P10	P9	P9	P9

TABLE VII
RANKING COMPARISON WITH INVERSION NUMBERS FOR RF CLASSIFICATION OF INDECS DATASET

Rank	Parameter	SIFT AE IN: 7	χ^2 IN: 22	Gain Ratio IN: 13	Info Gain IN: 19
1	P6	P1	P2	P5	P2
2	P1	P7	P5	P6	P5
3	P4	P5	P8	P7	P7
4	P7	P6	P7	P3	P6
5	P5	P2	P6	P2	P3
6	P3	P4	P3	P4	P8
7	P2	P3	P4	P8	P4
8	P8	P8	P1	P1	P1
9	P9	P9	P10	P10	P10
10	P10	P10	P9	P9	P9

TABLE VIII
RANKING COMPARISON WITH INVERSION NUMBERS FOR SVM CLASSIFICATION OF PASCAL 2006 DATASET

Rank	Parameter	SIFT AE IN: 26	χ^2 IN: 27	Gain Ratio IN: 14	Info Gain IN: 27
1	P7	P1	P1	P9	P1
2	P6	P4	P2	P1	P2
3	P9	P10	P5	P6	P5
4	P4	P6	P8	P7	P8
5	P5	P3	P7	P4	P7
6	P3	P5	P6	P3	P6
7	P2	P7	P3	P10	P3
8	P8	P2	P4	P5	P4
9	P1	P8	P10	P2	P10
10	P10	P9	P9	P8	P9

The results of the PASCAL 2006 dataset are shown in Table VIII. This dataset is the second most difficult among those we tested, as measured by the the final classification accuracy of our parameters. All of the methods had difficulty selecting useful parameters. Gain ratio achieves results comparable to the worst performing methods in the TRACE or medical datasets, and the other methods do much worse. All attribute evaluation methods rank the best performing parameter, relative smoothness (P7), as mediocre. They also all rank mean (P1) first or second, but it comes in ninth in the accuracy

ratings.

Given the universally poor performance, we suspect the primary failing here is not the parameter evaluation, but a combination of the difficulty of the dataset and the ability of the chosen parameters to represent it. It is possible that different parameters, more suited to PASCAL-style object recognition tasks, would be more amenable to our method on these datasets.

VIII. Conclusions

Across all of the datasets we studied, the classical parameter evaluation techniques were mostly unsuccessful in identifying the parameters best suited for correctly classifying the images with standard data mining classifiers. This lends support to the idea that an awareness of the spatial relationships between region extracted parameters is important for successful attribute selection in representations such as those produced by our grid-based segmentation.

For solar images, our SIFT AE method performed very well, beating the classical methods used in previous work by a substantial margin. If a reduction in parameter size is needed to accommodate space or computational limitations that arise in SDO CBIR as the mission's data accumulates, these results indicate that our method should be the best choice for determining which parameters to retain or drop to improve performance.

In addition to its success on solar images, the method performed comparably well on the medical datasets, which are known to be visually similar to the solar images; as well as the INDECS datasets, which were not previously considered similar. These results show that, for attribute selection on grid-extracted parameters, our method has clear applicability in multiple domains. Even on the most difficult datasets, the results of our method remained competitive with the performance of the other methods. With changes to the parameters used or other adjustments to better suit the specific properties of these domains, we expect that our method would surpass the classical methods here as well.

IX. Future Work

As the volume of solar data continues to grow, the need for scalable tools to accommodate research is essential. The immediate value of our work in integrating these attribute evaluation techniques into the SDO CBIR system as a way to improve on previous iterations of its design, helping it to more efficiently offer successfully image retrieval across the entirety of the eventual 5-10 year SDO data archive. Using our SIFT AE measure or further refinements to it should be able to near optimally reduce the parameters used to query the system, enabling near identical performance at a reduced computational cost. This, in addition to many other improvements planned, should make the SDO CBIR system a valuable solar research resource for years to come.

In addition to its use on solar data, we hope our technique proves useful in other astronomical applications, as well as in other domains. Strong parameter evaluation techniques should be valuable to the growing body of astronomical research utilizing image processing and massive astronomical data collections, such as the SuperCOSMOS Sky Survey [28] or the Galaxy Zoo Mergers project [29].

Outside astronomy, we plan to continue evaluating our method with other datasets, extending our tests to include other image parameters. This will allow us to better evaluate our method in a broader range of image retrieval environments. The use of different parameters better optimized for non-solar data may help us achieve strong performance in even more domains. We look forward to expanding the horizons of spatial parameter evaluation, and hope our work brings benefit not only to our own pursuits and to solar science research, but also many parts of the larger attribute evaluation community.

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. Martens, G. D. Attrill, A. Davey, A. Engell, S. Farid, P. Grigis, J. Kasper, K. Korreck, S. Saar, A. Savcheva *et al.*, "Computer vision for the solar dynamics observatory (sdo)," *Solar Physics*, vol. 275, no. 1, pp. 79–113, 2012.

[2] J. M. Banda and R. A. Angryk, "An experimental evaluation of popular image parameters for monochromatic solar image categorization," in *FLAIRS-23: Proceedings of the twenty-third international Florida Artificial Intelligence Research Society conference, Daytona Beach, Florida, USA*, 2010, pp. 380–385.

[3] M. A. Hall and G. Holmes, "Benchmarking attribute selection techniques for discrete class data mining," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 15, no. 6, pp. 1437–1447, 2003.

[4] L. C. Molina, L. Belanche, and À. Nebot, "Feature selection algorithms: A survey and experimental evaluation," in *Proceedings of the IEEE International Conference on Data Mining*. IEEE, 2002, pp. 306–313.

[5] "Solar dynamics observatory content-based image retrieval system," http://cbsir.cs.montana.edu/sdocbir/php/index.php, accessed: 2013-06-11.

[6] 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 *2010 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 2010, pp. 528–534.

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

[8] M. A. Schuh, J. M. Banda, P. N. Bernasconi, and R. A. Angryk, "A comparative evaluation of automated solar filament detection," unpublished.

[9] J. M. Banda, R. A. Angryk, and P. C. Martens, "On the surprisingly accurate transfer of image parameters between medical and solar images," in *18th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2011, pp. 3669–3672.

[10] H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medical applications-clinical benefits and future directions," *International journal of medical informatics*.

[11] R. C. Gonzalez and R. E. Woods, *Digital image processing*, 3rd ed. Prentice-Hall, Inc, 2008.

[12] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 8, no. 6, pp. 460–473, 1978.

[13] M. R. Schroeder, *Fractals, chaos, power laws: Minutes from an infinite paradise*. WH Freeman, New York, 1991.

[14] J. M. Banda, R. A. Angryk, and P. C. Martens, "On dimensionality reduction for indexing and retrieval of large-scale solar image data," *Solar Physics*, vol. 283, no. 1, pp. 113–141, 2013.

[15] 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 *IEEE International Conference on Fuzzy Systems, 2009. FUZZ-IEEE 2009*. IEEE, 2009, pp. 2019–2024.

[16] M. A. Schuh, R. A. Angryk, K. Gaensan Pillai, J. M. Banda, and P. C. H. Martens, "A large-scale solar image dataset with labeled event regions," in *Proc. of the International Conference on Image Processing (ICIP)*, 2013, 4349–4253.

[17] "Microsoft research cambridge object recognition image database," http://research.microsoft.com/en-us/downloads/b94de342-60dc-45d0-830b-9f6eff91b301/, accessed: 2013-06-11.

[18] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision*, vol. 88, no. 2, pp. 303–338, 2010.

[19] A. Pronobis and B. Caputo, "The kth-indecs database," *Tech. Rep. CVAP297, KTH Royal Institute of Technology*, 2005.

[20] W. Hersh, H. Müller, and J. Kalpathy-Cramer, "The imageclefmed medical image retrieval task test collection," *Journal of Digital Imaging*, vol. 22, no. 6, pp. 648–655, 2009.

[21] D. G. Lowe, "Object recognition from local scale-invariant features," in *The proceedings of the seventh IEEE international conference on Computer vision*, vol. 2. Ieee, 1999, pp. 1150–1157.

[22] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Workshop on statistical learning in computer vision, ECCV*, vol. 1, 2004, p. 22.

[23] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[24] J. R. Quinlan, "Induction of decision trees," *Machine learning*, vol. 1, no. 1, pp. 81–106, 1986.

[25] I. H. Witten, E. Frank, L. E. Trigg, M. A. Hall, G. Holmes, and S. J. Cunningham, "Weka: Practical machine learning tools and techniques with java implementations," *University of Waikato Research Commons*, 1999.

[26] A. Vedaldi and B. Fulkerson, "Vlfeat: An open and portable library of computer vision algorithms," in *Proceedings of the international conference on multimedia*. ACM, 2010, pp. 1469–1472.

[27] C. E. Leiserson, R. L. Rivest, C. Stein, and T. H. Cormen, *Introduction to algorithms*. The MIT press, 2001.

[28] N. Hambly, H. MacGillivray, M. Read, S. Tritton, E. Thomson, B. Kelly, D. Morgan, R. Smith, S. Driver, J. Williamson *et al.*, "The supercosmos sky survey–i. introduction and description," *Monthly Notices of the Royal Astronomical Society*, vol. 326, no. 4, pp. 1279–1294, 2001.

[29] C. J. Lintott, K. Schawinski, A. Slosar, K. Land, S. Bamford, D. Thomas, M. J. Raddick, R. C. Nichol, A. Szalay, D. Andreescu *et al.*, "Galaxy zoo: morphologies derived from visual inspection of galaxies from the sloan digital sky survey," *Monthly Notices of the Royal Astronomical Society*, vol. 389, no. 3, pp. 1179–1189, 2008.