ON THE SURPRISINGLY ACCURATE TRANSFER OF IMAGE PARAMETERS BETWEEN MEDICAL AND SOLAR IMAGES

Juan M. Banda [1], Rafal Angryk [1], Petrus Martens [2,3]

[1] Dept. Computer Science, Montana State University, Bozeman, MT, USA
[2] Dept. Physics, Montana State University, Bozeman, MT, USA
[3] Harvard-Smithsonian Center for Astrophysics, Cambridge, MA, USA

ABSTRACT

In this work we report on the transfer of image parameters that produce good results for medical images to the domain of solar image analysis. Using the first solar domain-specific benchmark dataset that contains multiple types of solar phenomena we discovered during our work for constructing a content-based image retrieval (CBIR) system for NASA's Solar Dynamics Observatory (SDO) mission that we could take advantage of the research on the creation of CBIR systems in the medical field. We demonstrate that, while image analysis is a very domain-specific task, there are lessons to be learned and methods to be shared between different fields. In this paper we present an extensive comparative analysis of several different domain-specific datasets in order to provide some guidance for the solar physics community on the well researched field of medical image processing allowing them to transfer knowledge from one applied field to their own.

***Index Terms*—** image processing, classification, content-based image retrieval, medical image retrieval

1. INTRODUCTION

In this work we present an investigation on how the solar physics community can be benefited by the knowledge that medical image researchers have been gathering over the years [1] to the task of image classification and retrieval in the area of Solar Physics. Research on automatic classification of solar images have been underway for several years and several missions, but usually has focused on specific types of solar phenomena [2, 3, 4]. With the recent launch of NASA's Solar Dynamics Observatory (SDO) mission, we have been presented with the unique opportunity of having a massive repository of high-quality solar images to analyze. This mission, with its Atmospheric Imaging Assembly (AIA), is expected to generate a set of eight 4096 pixels x 4096 pixels images every 10 seconds. This will lead to a data transmission rate of approximately 700 Gigabytes per day only from the AIA component (the entire mission is expected to be sending about 1.5 Terabytes of data per day), for a minimum of 5 years. With such and enormous volume of data, automated classification and retrieval becomes a daunting task. The lack of general research on solar image recognition for multi-class problems has made us turn to other domains to find knowledge we could transfer into our emerging solar domain.

One of the most important elements of developing a Content-Based Image Retrieval system for the SDO mission is selecting what image parameters to extract and store from the massive amounts of images that come down the SDO pipeline, and as any researcher in the image processing and retrieval fields should know, this is not a trivial task due to the domain-specific dependency of the image parameters. All of the SDO solar images are grayscale and this allowed us to look into image parameters that performed well for these types of images, and as we found out in [5] and [6] there had been quite extensive amounts of work done for the classification and annotation of medical radiograph images with certain textural image parameters. As we incorporated some of these parameters we also experimented with several others that have worked in this domain but were deemed too expensive or not ideal enough for our particular images [7-9]. We mainly took advantage of the analysis performed to build medical CBIR systems in the past and present [5, 6]. However, none of these successful CBIR systems have dealt with the volume of data that the SDO mission will generate, making our task of avoiding redundant and less useful image parameters a very critical one for the success of our SDO CBIR system.

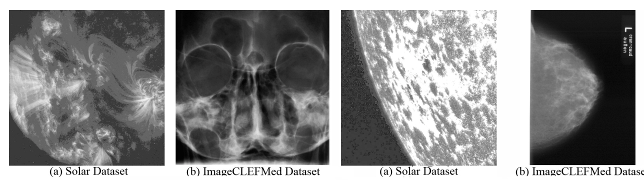


Figure 1. Sample images of the Solar dataset [9] (a) and the ImageCLEFMed medical dataset [10] (b)

The purpose of this work is to show direct applicability of research on medical radiograph image analysis and retrieval to the research conducted on solar images. By publishing this paper we want to make the solar community aware of the benefits of transferring findings from the medical domain and encourage them to apply these findings into solar image analysis in the future. There is plenty to be learned and transferred from other domains with images that have certain similarities (i.e. medical radiograph images, see fig. 1). However, there are still considerable amounts of experimentation and refining to be done once one has determined a domain that can provide valuable contributions and strong foundations to the particular problem being worked on, as we have learned over the course of our trek in building a CBIR system [7-9].

In order to prove our point, we present a comparative evaluation of the performance of our selected image parameters for six different datasets and validate our own selection of image parameters for our domain-specific solar dataset. We also demonstrate how while we can still transfer some things from one domain to another, there are particular domain-specific datasets where our parameters will fail to produce any useful results in terms of image classification.

The rest of the paper is organized in the following way: Sec. 2 the image parameters utilized, Sec. 3 includes a description of the datasets we used in our experiments. In Sec. 4 we present the classifiers used for our comparative evaluation. Our experimental results can be found in section 5, and our conclusions are listed in section 6.

2. IMAGE PARAMETERS

Based on our literature reviews, we decided that we would use some of the most popular image parameters used in different fields such as medical images, text recognition, natural scene images and traffic images [5, 11-14], as a common denominator the usefulness of all these image parameters have shown to be very domain dependent. We used grid based image segmentation with 64 cells per image (i.e. 8 x 8 grid), since the image segmentation technique has been proven to produce good results [9, 15]. We calculated all 10 features for each individual cell.

The ten image parameters [16] that we have concluded to be the most useful for our task are: the Mean intensity, the Standard Deviation of the intensity, the Third Moment and Fourth Moment, Uniformity, Entropy, and Relative Smoothness, Fractal Dimension [17], and two Tamura texture attributes: Contrast and Directionality [18].

The values produced by our texture feature extraction formulas are not influenced by different orientations of the same kinds of phenomena in different images [19], and can be extracted very quickly [20].

3. BENCHMARK DATASETS

In order to better compare and contrast the image parameters selected, we have opted to experiment on 6 different dataset. Based on our hypothesis, we believe ImageClefMed[10] 2005 and 2007 are the most similar to our dataset while the INDECS[21] and PASCAL[22] 2006 and 2008 datasets provide completely different images that will nicely outline the importance of the selection of the image parameters and the domain-specific nature of said parameters. These datasets were chosen due to their popularity in their individual domains.

3.1. Solar Dataset

The dataset, first introduced in [9] consists of 1,600 images divided in 8 equally balanced classes representing 8 types of different solar phenomena: *Active Region, Coronal Jet, Emerging Flux, Filament, Filament Activation, Filament Eruption*, Flare and Oscillation. All of our images are 1,024 by 1,024 pixels.

The benchmark dataset both in its original and pre-processed format is freely available to the public via Montana State University's server [23]. These images where taken from the TRACE mission, since at the time our work started the SDO mission was not launched and TRACE data was what other modules had been using for testing and developing their algorithms.

3.2. Medical Datasets: ImageCLEFMed 2005 and 2007

ImageCLEF is the cross-language image retrieval track which is run as part of the Cross Language Evaluation Forum (CLEF) [24]. ImageCLEF has participation from both academic and commercial research groups. In their medical retrieval task, we find several datasets available from 2005 to 2010.

The 2005 dataset consists of 10,000 radio graphs categorized in 57 different categories. For the 2007 dataset, the number of categories doubled to 116 and the number of images increased by two thousand. This dataset compiles anonymous radiographs, which have been arbitrarily selected from routine at the Department of Diagnostic Radiology, Aachen University of Technology (RWTH). The images in this dataset present different ages, genders, view positions and pathologies. We randomly selected 8 different classes from the 2005 and 2007 datasets creating two balanced datasets with 1,600 images each.

3.3. Non-Medical Dataset: INDECS

INDECS is an acronym for Indoor Environment under Changing conditionS. This dataset consists of several sets of pictures taken in five rooms of different functionality under various illumination and weather conditions at different periods of time. After selecting 8 classes from this dataset we balanced the classes by selecting at random 200 images from each of them.

This dataset features very different images as the ones found in the TRACE and can be obtained freely on the web at [25], both in high and low resolution quality.

3.4. Non-Medical Datasets: PASCAL 2006 and 2008

The PASCAL Visual Object Classes (VOC) challenge is a popular benchmark in visual object category recognition and detection, providing the vision and machine learning communities with a standard dataset of images, annotation, and standard evaluation procedures. This dataset has been gaining popularity in the community since its conception in 2005. The main challenges of the PASCAL challenge are classification and detection of several different classes of objects within their images, we selected this dataset for their classification challenge and to observe how our domain-specific image parameters behave.

The PASCAL 2006 dataset consists of 10 classes, from which we selected 8 classes at random for our comparative analysis. For the PASCAL 2008 dataset of 20 classes, we again selected 8 different classes. For both datasets we selected 200 random images from each class.

4. COMPARATIVE ANALYSIS BASED ON CLASSIFIERS

We selected Naïve Bayes and Support Vector Machines (SVM) with a linear kernel function as our linear classifiers and C4.5 as a decision tree classifier. Linear classifiers achieve the grouping of items that have similar feature values into groups by making a classification decision based on the value of the linear combination of the features. Whereas C4.5 uses entropy-based information gain measure to split samples into classes.

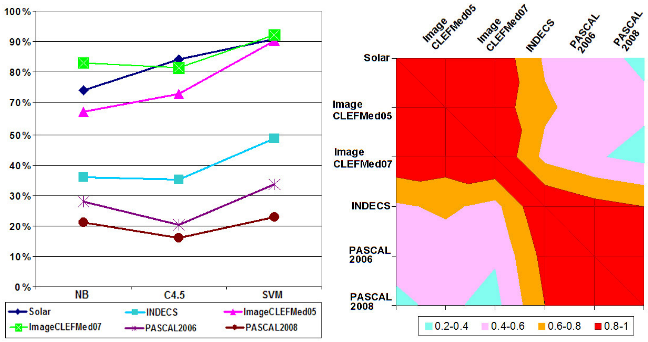
The Naïve Bayes classifier has shown to be surprisingly accurate in most cases [26] and executes very fast, making it a highly relevant candidate for training on the massive number of images expected in the SDO repository. C4.5 being part of the decision tree classifiers is one of the most popular in the community, takes a greedy approach and is also quick to compute. Finally, the SVM classifiers have gained popularity in the scientific data mining community in the last few years due to their ability of maximizing separation functions which improve the classifier accuracy and ability of extrapolating results on new unseen data with large datasets [27]. Our main concern for applying SVM's to new SDO data is their slow learning process but we include them in our comparative analysis.

Note that we use these classifiers in order to present a comparative analysis of our experiments between the datasets. In this paper we are not trying to find the best classification results or tweak the classifiers to perform at its best.

Besides the classification accuracy percentages, we present the Receiver Operating Characteristic (ROC) curves for the best performing classifiers in order to observe how precise our accuracy is for certain classes within the datasets and to compare individual datasets on a per-class basis. For the members of the solar image processing community we include the precision, recall and F-measure values for each class in our experimental results on the solar data as an appendix at the end of this paper.

5. EXPERIMENTAL RESULTS

All classification experiments have been run using 10-fold cross validation. In figure 2 we present the classification accuracy charts for all six different datasets we investigated on.



(a) (b)

Figure 2. (a) Percentage of classification accuracy for our three different classifiers on the six investigated datasets. (b) Correlation heat map of the classification accuracy between datasets

As it can be observed in fig, 2(a) and (b), we have two main clusters when it comes to classification results. On the top cluster of fig. 2(a) ranging between 74% and 92% classification accuracy, we have the three somewhat similar datasets: Solar, ImageCLEFMed05 and ImageCLEFMed06. On the bottom cluster we have the three datasets from other domains: Pascal 2006, 2008 and INDECS. Is clearly evident that we can transfer some of the most common image parameters used for medical images to our solar images and achieve very acceptable classification results. In figure 2(b) we can clearly observe the clusters of correlations between datasets. This figure shows that the correlation between their classification accuracy for the solar and medical datasets is very high. It also contrasts them from the non-medical datasets, and shows the high correlation of low results between them.

In the figure 3 we present the ROC curves for all 8 classes in the 6 different datasets for the best performing classifier (SVM). The second measure (ROC Curve) follows the same trends than on figure 2. All though our SVM had 'hard' time recognizing one of the classes in the Solar (a) and ImageCLEFMed07 (c) datasets, we still have a very clear separation between the solar and medical datasets (a, b, c) and the non-medical ones (d, e, f). In figure 3 the ROC curves show good (over 0.8) to excellent (over 0.9) results for almost all classes in the medical and solar datasets, as for the PASCAL and INDECS datasets we see them dropping below 0.6, for the majority of the classes, making the recognition results very bad (random selection generates 0.5 area) for the used image parameters. All our results are easy to replicate and you can find all the files used to generate these results in the supplemental website [31].

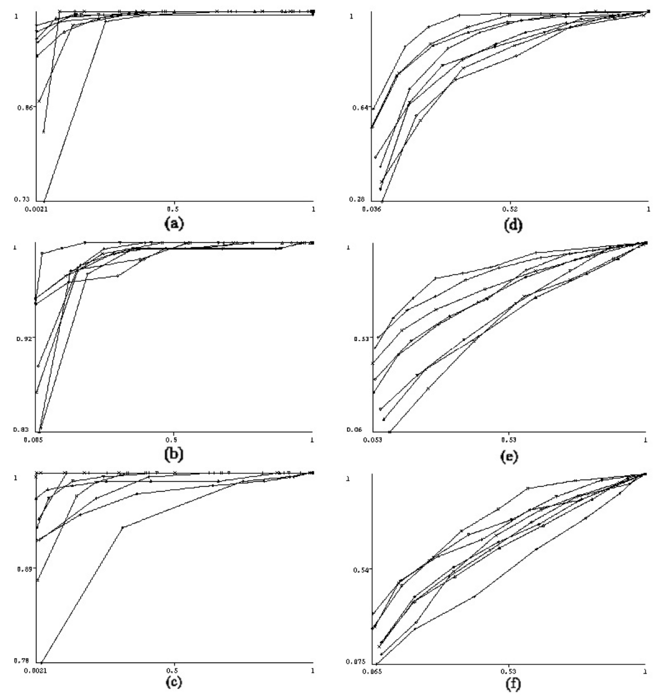


Figure 3. ROC Curves of individual class labels (eight for each dataset) for the SVM Classifier. (a) Solar Dataset, (b) ImageCLEFMed05 dataset, (c) ImageCLEFMed07 dataset, (d) INDECS dataset, (e) PASCAL 2006 dataset, (f) PASCAL 2008 dataset

6. CONCLUSIONS

The experiments from section 5 clearly demonstrate that some image parameters while domain-dependent can still be applied to other domains where the images being analyzed are highly similar. This is clearly shown when comparing results for solar and medical data (ImageCLEFMed05 and 07, see figure 2). They both achieve similar classification accuracy percentages due to the fact that our textural image parameters are similarly effective for the categorization of solar phenomena and radiograph medical images. The characteristics of both of these types of images have factors in common, such as image mode (grayscale), fuzzy edges and the objects in them benefiting from a grid based parameter extraction approach. Creating a feature vector in a similar fashion for both has proven to produce good results as presented by for the medical images [5, 6] and for solar images [7-9]. In this paper we show clearly by making side-by-side comparisons that the ROC curves in figure 3 complement our findings by showing good to excellent results for the medical and solar datasets for almost all classes.

In contrast to the medical and solar results, some parameters on the INDECS and PASCAL 2006 and 2008, lead to very bad results because of the nature of the images. These domains contain images with clear edges, regular shapes and feature high levels of dissimilarity between them, only having certain objects in common. As suggested by [5], textural image parameters tend to perform poorly on images of this kind, and we clearly confirmed that in our experiments. Researchers focusing on analyzing these types of images have turned to different types of image parameters such as SIFT descriptors [28], color histograms [29], and color patches [30], that produce better results for different domains. While we can continue analyzing image parameters for years, we feel that based on our research and the work done in the medical field, the solar research community should take under consideration utilizing some of the most informative and useful parameters for works on medical data.

This work goes to shows that while image processing is a very domain-specific task that highly depends on the types of images being analyzed, there are some fields that can help each other and that some research can be transferred into the solar image field.

7. ACKNOWLEDGEMENTS

This work was supported in part by the NASA Grant Award No. 08-SDOSC08-0008, funded from NNH08ZDA001N-SDOSC: Solar Dynamics Observatory Science Center solicitation.

8. APPENDIX

Table 1. Accuracy of the SVM classifier for the individual groups of Solar phenomena

Class	Precision	Recall	F-Measure
Active Region	0.975	0.980	0.978
Coronal Jet	0.941	0.950	0.945
Emerging Flux	0.980	0.970	0.975
Filament	0.915	0.965	0.939
Filament Activation	0.975	0.960	0.967
Filament Eruption	0.903	0.880	0.891
Flare	0.770	0.735	0.752
Oscillation	0.805	0.825	0.815
Average	**0.908**	**0.908**	**0.908**

9. REFERENCES

[1] H. Müller, N. Michoux, et al. "A review of content-based image retrieval systems in medical applications--clinical benefits and future directions." International journal of medical informatics 73(1): 1-23. 2004

[2] V. Zharkova, S. Ipson, A. Benkhalil, and S. Zharkov. "Feature recognition in solar images". Artificial Intelligence Review, 23(3), pp. 209–266. 2005

[3] S. Christe, I.G. Hannah, S. Krucker, J. McTiernan, and R.P. Lin "RHESSI Microflare Statistics. I. Flare-Finding and Frequency Distributions". ApJ, 677, pp. 1385–1394. April 2008

[4] P.N. Bernasconi, D.M Rust, and D. Hakim. "Advanced automated solar filament detection and characterization code: description, performance, and results." Sol. Phys., 228 pp. 97–117. May 2005

[5] T. Deselaers, D. Keysers, and H. Ney. Features for Image Retrieval: An Experimental Comparison. Information Retrieval, vol. 11, issue 2, Springer, 77-107. The Netherlands 03/2008

[6] H. Müller, A. Rosset, et al. "A reference data set for the evaluation of medical image retrieval systems." Computerized Medical Imaging and Graphics 28(6): 295-305. 2004

[7] J.M Banda and R. Angryk "Usage of dissimilarity measures and multidimensional scaling for large scale solar data analysis". Proceedings of the 2010 Conference on Intelligent Data Understanding (CIDU) 2010, Mountain View, California, USA. pp 189-203. October 5-6, 2010

[8] J.M Banda and R. Angryk "Selection of Image Parameters as the First Step Towards Creating a CBIR System for the Solar Dynamics Observatory". International Conference on Digital Image Computing: Techniques and Applications (DICTA). Sydney, Australia, December 1-3, 2010.

[9] J.M Banda and R. Angryk "An Experimental Evaluation of Popular Image Parameters for Monochromatic Solar Image categorization" Proceedings of the 23rd international Florida Artificial Intelligence Research Society conference (FLAIRS-23), Daytona Beach, Florida, USA. pp. 380-385. May 19–21 2010

[10] W. Hersh, H. Müller, J. Kalpathy-Cramer, E. Kim, X. Zhou, "The consolidated ImageCLEFmed Medical Image Retrieval Task Test Collection", Journal of Digital Imaging, volume 22(6), 2009, pp 648-655.

[11] S. Kulkarni, B. Verma, "Fuzzy Logic Based Texture Queries for CBIR," Fifth International Conference on Computational Intelligence and Multimedia Applications (ICCIMA'03), pp.223, 2003

[12] H Lin, C Chiu, and S. Yang, "LinStar texture: a fuzzy logic CBIR system for textures", In Proceedings of the Ninth ACM international Conference on Multimedia (Ottawa, Canada). MULTIMEDIA '01, vol. 9. ACM, New York, NY, pp 499-501. 2001.

[13] S. Thumfart, W. Heidl, J. Scharinger, and C. Eitzinger. "A Quantitative Evaluation of Texture Feature Robustness and Interpolation Behaviour". In Proceedings of the 13th international Conference on Computer Analysis of Images and Patterns. 2009.

[14] J. Muwei, L. Lei, G. Feng, "Texture Image Classification Using Perceptual Texture Features and Gabor Wavelet Features," Asia-Pacific Conference on Information Processing vol. 2, pp.55-58, 2009.

[15] J.M Banda and R. Angryk "On the effectiveness of fuzzy clustering as a data discretization technique for large-scale classification of solar images" Proceedings of the 18th IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '09), Jeju Island, Korea, pp. 2019-2024 August 2009

[16] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten. "The WEKA Data Mining Software: An Update" SIGKDD Explorations, Volume 11, Issue 1, 2009

[17] M. Schroeder. Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise. New York: W. H. Freeman, pp. 41-45, 1991.

[18] H. Tamura, S. Mori, T. Yamawaki. "Textural Features Corresponding to Visual Perception". IEEE Transaction on Systems, Man, and Cybernetics 8(6): pp. 460–472. 1978.

[19] R.M Haralick, K. Shanmugam and I. Dinstein, "Textural Features For Image Classification," IEEE Transactions on Systems, Man, and Cybernetics, Volume: SMC-3, No. 6, pp 610- 621. 1978.

[20] R. Lamb, "An Information Retrieval System For Images From The Trace Satellite," M.S. thesis, Dept. Comp. Sci., Montana State Univ., Bozeman, MT, 2008.

[21] A. Pronobis, B. Caputo, P. Jensfelt, and H. I. Christensen. "A discriminative approach to robust visual place recognition". In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS06), Beijing, China, 2006.

[22] M. Everingham, L. Van Gool, C.K.I. Williams, J. Winn, and A. Zisserman. "The Pascal Visual Object Classes (VOC) Challenge". International Journal of Computer Vision, 88(2), 303-338, 2010

[23] SDO Dataset (MSU) [Online], Available: http://www.cs.montana.edu/angryk/SDO/data/ [Accessed: Jan 10, 2011]

[24] Cross Language Evaluation Forum [Online], Available: http://www.clef-campaign.org/ [Accessed: Jan 10, 2011]

[25] The INDECS Database [Online], Available: http://cogvis.nada.kth.se/INDECS/ [Accessed: Jan 10, 2011]

[26] D. Hand, and K. Yu. "Idiot's Bayes-not so stupid after all?" International Statistical Review 69(3): 385-398. 2001

[27] I.Tsang, J. Kwok, et al. "Core vector machines: Fast SVM training on very large data sets." Journal of Machine Learning Research 6(1): 363. 2006

[28] D.G. Lowe. "Object recognition from local scale-invariant features". Proceedings of the International Conference on Computer Vision. 2. pp. 1150–1157. 1999

[29] A. Jain, and A. Vailaya. "Image retrieval using color and shape." Pattern Recognition 29(8): 1233-1244. 1996

[30] T. Gevers, and A. Smeulders. "Pictoseek: Combining color and shape invariant features for image retrieval." Image Processing, IEEE Transactions on 9(1): 102-119. 2002

[31] ICIP 2011 Supplemental Website [Online], Available: http://www.jmbanda.com/ICIP2011/ [Accessed: Jan 20, 2011]