Skin Lesion Image Recognition with Computer Vision and Human in the Loop

Orod Razeghi[1]
http://www.cs.nott.ac.uk/~ozr
Guoping Qiu[1]
http://www.cs.nott.ac.uk/~qiu
Hywel Williams[2]
Hywel.Williams@nottingham.ac.uk
Kim Thomas[2]
Kim.Thomas@nottingham.ac.uk

[1] VIPLAB, IMA Group
Computer Science
University of Nottingham
Nottingham, UK

[2] Centre of
Evidence-Based Dermatology
University of Nottingham
Nottingham, UK

Abstract

In this paper, we present an interactive human in the loop computer vision technique for the recognition of skin lesion images. We have designed a dermatology "Question and Answer" bank suitable for interactively extracting human perceptual knowledge of images in order to assist computer vision algorithms in boosting recognition accuracies. We present experimental results to show that for some diseases, traditional computer vision techniques can only achieve a recognition rate of 20%, whilst with human in the loop the performance can be boosted to over 96%. We also show that users do not require any medical knowledge to answer these questions to achieve excellent recognition rates.

1 Introduction

Recently, researchers have advocated and developed the so-called interactive imaging and vision [11] or human in the loop [4] approach to visual object recognition. It has been well known that some problems, that are difficult for computer to solve, are actually very simple for human. For instance, fully automated and accurate computer visual object recognition has proved to be very difficult, if not entirely impossible. An intermediate solution to develop useful and practical technology is to make human and computer work in harmony and exploit their respective strengths. We believe that the approach introduced in [4] is particularly suitable for medical image recognition applications, such as computer aided diagnosis.

Unlike the majority of publications in the area of computer vision for dermatology applications, this paper is the first that attempts to apply promising human in the loop visual recognition technique to automatic recognition of various conditions, including non-melanoma skin diseases, and is certainly one that achieves the best results in the literature. Our contributions include: i) demonstrated for the first time that human in the loop visual recognition can significantly boost accuracy and achieve near perfect recognition results, ii) designed a dermatology relevant "Q&A" bank containing 21 questions and over 100 possible answers suitable for human in the loop visual recognition solutions and iii) shown that such system can be exploited by users without any medical knowledge.

2 Computer Vision in Dermatological Applications

An example of related work is an image analysis system presented in [1] that differentiates early melanoma from benign pigmented lesions. The analysis system extracts features related to the size, shape, boundary, and colour of each lesion. Another solution in form of an automated melanoma recognition system is presented in [6]. A binary mask of lesion is obtained by a number of basic segmentation algorithms alongside a fusion strategy. A set of shape and radiometric features are calculated to determine the malignancy of a lesion.

A model of tissue colouration is presented in [5]. The model is built by computing the spectral composition of light remitted from normal human skin colour, and comparing it to abnormal tissues. As an alternative the framework in [12] assesses a series of 588 flat pigmented skin lesions. The proposed analyser employs an artificial neural network. A feature selection procedure confirms that as few as 13 variables are adequate to discriminate the two groups of "melanoma" and "other pigmented" skin lesions.

Although the literature demonstrates a number of attempts at fabricating Content Based Image Retrieval (CBIR) Medical Systems for dermatological purposes [3][7], and quite a few attempts at assessing severity of specific skin diseases automatically [13], the lack of a reliable system for unskilled users, or an assistant tool for general practitioners is apparent.

3 Human in the Loop Skin Lesion Recognition

Our system adopts the framework of [4] for incorporating any multi-class object recognition algorithm that produces a probabilistic output over classes, as follows:

$$p\ c\ x\ U \qquad \frac{p\ U\ c\ x\ p\ c\ x}{{}_c p\ U\ c\ p\ c\ x} \qquad \frac{p\ U\ c\ p\ c\ x}{{}_c p\ U\ c\ p\ c\ x} \tag{1}$$

where c is class, x is image, and U is any random sequence of user answers. The assumption that $p\ U\ c\ x \quad p\ U\ c$ suggests that the types of randomness present in user answers is class-dependent and not image-dependent.

In our implementation of the above framework, we employed 10 image features including Coloured Pattern Appearance Model (CPAM) [10], Geometric Blur (GB) [15], Global Image Descriptor (GIST) [8], Pyramid Histogram of Oriented Gradients (PHOG) and its variations [15], Scale-invariant Feature Transform (SIFT) and its variations, Pyramid Histogram of Visual Words (PHOW) and its variations [14], and Self-similarity Feature (SSIM) [15]. We used *OBSCURE* [9], a state of the art, publicly available open source Support Vector Machine (SVM) based classifier for the visual classification, $p\ c\ x$.

Similar to the original work [4], we also used a multinomial distribution with a Dirichlet prior to model user responses $p\ U\ c$, and used KL divergence and maximum information gain to choose the next suitable question. Figure 1 illustrates the entire process.

4 Experimental Results

4.1 Datasets

We have collected two datasets from various Internet sources. The first and second datasets contain 90 and 706 dermatological images from 3 and 7 different skin diseases respectively. The lesions were manually segmented using a bounding box that includes pixels of lesion,

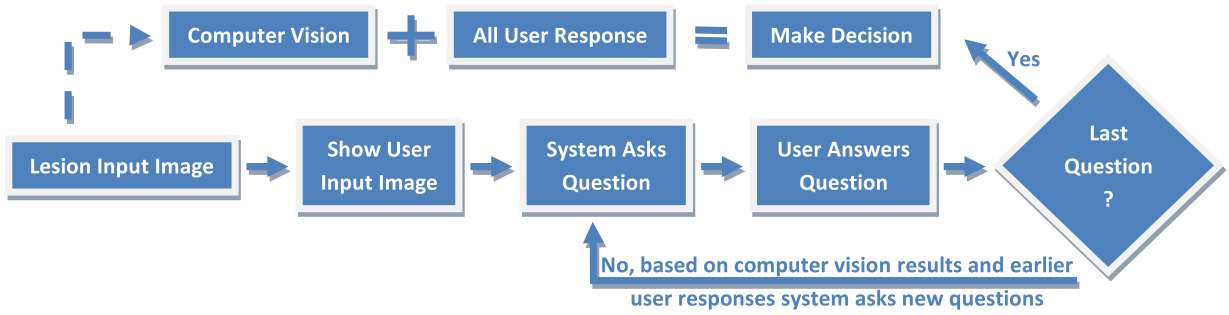

Figure 1: A skin lesion image is displayed to the user. For each image, a question and its possible answers are also displayed. The user answers the question. The user will repeat the process until all the questions are finished or a threshold is reached. The next question is selected by looking at previous user's answers and computer vision input. The final decision is also made by combining these two elements.

healthy skin and noise, such as hair. Features were extracted from the entire bounding box, which as a whole is treated as a single instance. Images with their ground truth classification were mainly collected from http://www.dermis.net.

4.2 Dermatology Question and Answer Bank

A set of questions, which both summaries patient general conditions as well as her skin lesion characteristics, were designed to help with obtaining user perception of patient in a (fabricated) scenario. The questions range from the age of patient to colour and surface features of skin lesions. We have consulted medical professionals and a dermatological reference [2] to scientifically derive these questions.

Table 1 lists 8 questions and 36 possible answers used for testing the first dataset and Table 2 lists 13 questions and 67 possible answers used for testing the second dataset. Wherever specific medical terms are used, a guide image with explanations is available for users to avoid confusion.

4.3 Results

Table 3 shows the results for the 1st and 2nd datasets. In the 1st dataset, 15 randomly selected images from each of the 3 diseases were used in training and the rest were used for testing.

Table 1: Dermatology First Dataset Questions

Question	Attributes (Possible Answers)
01 Site	Head, Trunk, Arms, Legs
02 Condition	Acute, Chronic
03 Surface	Normal, Scaly, Hyperkeratotic, Warty, Crust, Exudate, Excoriated
04 Lesion	Flat, Raised, Fluid Filled, Surface Broken
04 Colour	Pink, Red, Purple, Mauve, Brown, Black, Blue, White, Yellow, ...
06 Age	Infant, Young, Adult, Old
07 Contagiousness	Contagious, Non-contagious
08 Itchiness	Itchy, Non-itchy

Table 2: Dermatology Second Dataset Questions

Question	Attributes (Possible Answers)
01 Age	Infant, Child, Adult, Elderly
02 History	Personal, Family
03 Site	Face, Scalp, Ears, (Mouth, Tongue, Lips), Trunk, Hands, ...
04 Number	Single, Multiple
05 Distribution	Symmetrical, Asymmetrical, Unilateral, Localised, Generalised
06 Arrangement	Discrete, Coalescing, Disseminated, Annular, Linear, Grouped
07 Erythema	Erythematous, Non-erythematous
08 Duration	Acute, Chronic
09 Type	Flat, Raised Solid, Fluid Filled, Cyst, Comedone, Broken Surface
10 Surface	Normal, Abnormal Keratinisation, Scale, Broken, Crust, Shiny, ...
11 Colour	Due to blood (Red, Pink), Due to pigment (Black, Blue), ...
12 Border	Well defined, Poorly defined, Accentuated edge
13 Shape	Round, Irregular, Rectangular, Serpiginous, Dome shaped, ...

The experiment was repeated 5 times and the results in the table is the average over 5 rounds of experiments by a group of non-expert users. Here, it is clear that computer vision performs very badly on the Scabies images, achieving only 33% correct recognition rate. With human in the loop, the correct recognition rate boosts to 93% - a very significant improvement. Average correct recognition rate across the diseases is just over 57% for computer vision only solution but it is boosted to over 97% with human in the loop.

In the 2nd dataset, 30 images from each disease were randomly selected for training and the remaining 496 images were used for testing. Here the computer vision technique can only achieve 20% recognition rate for Mycosis Fungoides, but with human in the loop, the recognition rate is boosted to over 96% - again a dramatic improvement. The average across the diseases for the entire dataset is 61% for computer vision only and 96% for introducing human in the loop.

These results clearly demonstrate the effectiveness of human in the loop technique for recognising skin lesions. Compared to computer vision only solutions, adding human in the loop can dramatically improve the correct recognition rates. Moreover, computer aided diagnosis is known to be capable of reducing subjectivity, thus can reduce inter observer discrepancies, and our highly accurate results are consistent with this conclusion.

It may be argued that contribution of computer vision is unclear, since human responses are adequate to correctly classify a disease. However, as shown in figure 2, computer vision plays an important role in reducing human labour in terms of number of questions necessary to answer, and the time spent on each image to correctly classify a disease. Furthermore, some images cannot be classified correctly without computer vision, even after asking all the questions. There are over 1000 skin conditions worldwide, a fully functioning system will need a question bank of hundreds, if not thousands of questions. Computer vision will be essential in reducing labour and improving diagnosis.

More interestingly, it was also found that users do not have to answer all the questions correctly in order to achieve correct recognition. It was seen that although users' questions are asked in different orders and users' answers to the same questions are different, the algorithm still recognises images successfully. Figure 3 illustrates the frequency of possible answers selected by users in the 2nd dataset.

Table 3: 1st and 2nd Dataset Accuracy Results

Class (1st Dataset)	Visual	Vis+Attr
Discoid Eczema	80%	100%
Infantile Acne	60%	100%
Scabies	33.33%	93%
Total	57.78%	97.66%
Class (2nd Dataset)	**Visual**	**Vis+Attr**
Allergic Vasculitis	59.52%	100%
Atopic Eczema	48.67%	89.38%
Bullous Pemphigoid	56.09%	100%
Lichen Planus	38.09%	95.23%
Mycosis Fungoides	20.75%	96.22%
S. Cell Carcinoma	67.74%	98.38%
S. S. Melanoma	91.6%	98.6%
Total	61.09%	96.16%

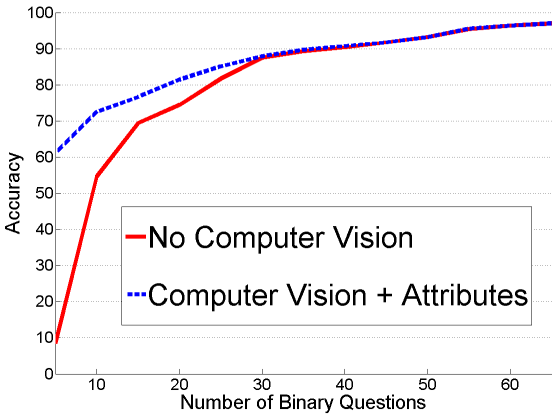

Figure 2: Results when no CV is involved. Although, CV and CV+Attributes results merge after a number of questions, CV increases accuracy without troubling the user with too many questions.

5 Conclusions

It is believed that there are between 1000 to 2000 skin conditions, and about 20% are difficult to diagnose. In the UK typical general practitioners receive minimal dermatology training. Our promising results from non-medical experts illustrate the potential clinical application of our work for health care providers, and also for places where access to health services are scarce.

We believe that we have for the first time applied a human in the loop visual recognition technique to diagnosis of skin diseases from visual images of affected areas. We have shown for some of the conditions, computer vision technique performs very poorly (as low as 20%), whilst human in the loop technique boosts the recognition rate to over 96%. Our future work is to apply the technique to a larger number of diseases, to refine the "Question and Answer" bank, and to implement the work on a smart mobile phone.

Acknowledgement: This work is partially supported by an EPSRC grant (EP/J501499/1).

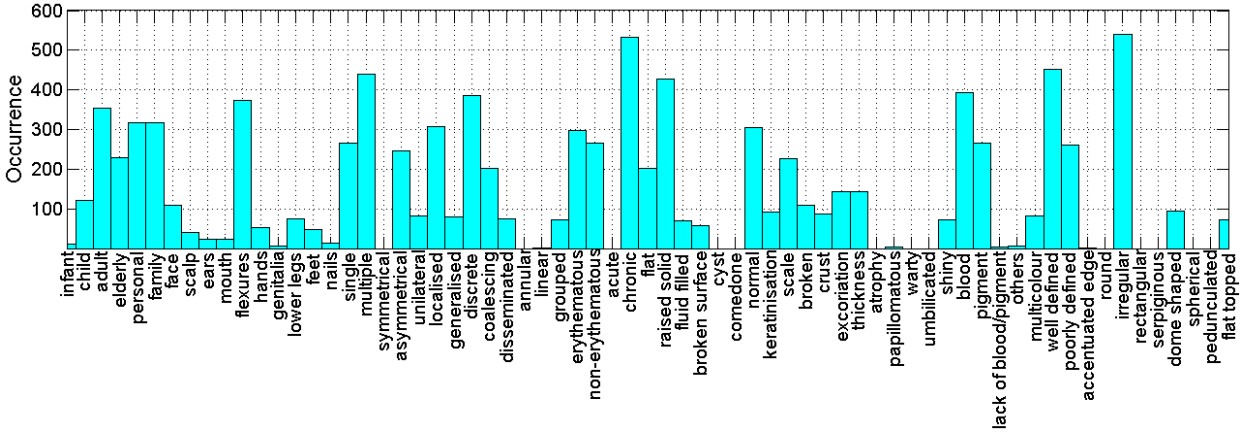

Figure 3: Frequency of answers picked by users in the 2nd dataset

References

[1] J Pfitzner M O'Rourke N Knight A Green, N Martin. Computer image analysis in the diagnosis of melanoma. *Journal of the American Academy of Dermatology*, 1994.

[2] R. Ashton and B. Leppard. *Differential diagnosis in dermatology*. Radcliffe, 2005.

[3] Lucia Ballerini, Xiang Li, Robert B Fisher, Ben Aldridge, and Jonathan Rees. Content-based image retrieval of skin lesions by evolutionary feature synthesis. *Applications of Evolutionary Computation*, 2010.

[4] Steve Branson, Catherine Wah, Florian Schroff, Boris Babenko, Peter Welinder, Pietro Perona, and Serge Belongie. Visual recognition with humans in the loop. In *Proceedings of the 11th European conference on Computer vision: Part IV*.

[5] Ela Claridge, Symon Cotton, Per Hall, and Marc Moncrieff. From colour to tissue histology: Physics based interpretation of images of pigmented skin lesions. In *Medical Image Computing and Computer-Assisted Intervention MICCAI 2002*.

[6] H. Ganster, P. Pinz, R. Rohrer, E. Wildling, M. Binder, and H. Kittler. Automated melanoma recognition. *Medical Imaging, IEEE Transactions on*, 2001.

[7] Henning Muller, Antoine Rosset, Jean-Paul Vallee, and Antoine Geissbuhler. Integrating content-based visual access methods into a medical case database. In *Proceedings of the Medical Informatics Europe Conference (MIE 2003)*.

[8] Aude Oliva and Antonio Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int. J. Comput. Vision*, 2001.

[9] Francesco Orabona, Jie Luo, and Barbara Caputo. Online-batch strongly convex multi kernel learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

[10] G. Qiu. Indexing chromatic and achromatic patterns for content-based colour image retrieval. *Pattern Recognition*, 2002.

[11] Guoping Qiu and Pong C. Yuen. Editorial: Interactive imaging and vision-ideas, algorithms and applications. *Pattern Recognition*, 2010.

[12] Pietro Rubegni, Gabriele Cevenini, Marco Burroni, Roberto Perotti, Giordana Dell'Eva, Paolo Sbano, Clelia Miracco, Pietro Luzi, Piero Tosi, Paolo Barbini, and Lucio Andreassi. Automated diagnosis of pigmented skin lesions. *International Journal of Cancer*, 2002.

[13] L Savolainen, J Kontinen, E Alatalo, J Rning, and A Oikarinen. Comparison of actual psoriasis surface area and the psoriasis area and severity index by the human eye and machine vision methods in following the treatment of psoriasis. *Acta dermatovenereologica*, 1998.

[14] A. Vedaldi and B. Fulkerson. Vlfeat: An open and portable library of computer vision algorithms. http://www.vlfeat.org/, 2008.

[15] A. Vedaldi, V. Gulshan, M. Varma, and A. Zisserman. Multiple kernels for object detection. In *Proceedings of the International Conference on Computer Vision (ICCV)*.