

# A Query-by-Example Content-Based Image Retrieval System of Non-Melanoma Skin Lesions

Lucia Ballerini<sup>1</sup>, Xiang Li<sup>1</sup>, Robert B. Fisher<sup>1</sup>, and Jonathan Rees<sup>2</sup>

<sup>1</sup> School of Informatics, University of Edinburgh, UK  
x.li-29@sms.ed.ac.uk, lucia.ballerini@ed.ac.uk, rbf@inf.ed.ac.uk

<sup>2</sup> Dermatology, University of Edinburgh, UK  
jonathan.rees@ed.ac.uk

**Abstract.** This paper proposes a content-based image retrieval system for skin lesion images as a diagnostic aid. The aim is to support decision making by retrieving and displaying relevant past cases visually similar to the one under examination. Skin lesions of five common classes, including two non-melanoma cancer types are used. Colour and texture features are extracted from lesions. Feature selection is achieved by optimising a similarity matching function. Experiments on our database of 208 images are performed and results evaluated.

## 1 Introduction

Research in content-based image retrieval (CBIR) today is an extremely active discipline. There are already review articles containing references to a large number of systems and description of the technology implemented [1, 2]. A more recent review [3] reports a tremendous growth in publications on this topic. Applications of CBIR systems to medical domains already exist [4], although most of the systems currently available are based on radiological images.

Most of the work in dermatology has focused on skin cancer detection. Different techniques for segmentation, feature extraction and classification have been reported by several authors. Concerning segmentation, Celebi et al. [5] presented a systematic overview of recent border detection methods: clustering followed by active contours are the most popular. Numerous features have been extracted from skin images, including shape, colour, texture and border properties [6–8]. Classification methods range from discriminant analysis to neural networks and support vector machines [9–11]. These methods are mainly developed for images acquired by epiluminescence microscopy (ELM or dermoscopy) and they focus on melanoma, which is actually a rather rare, but quite dangerous, condition whereas other skin cancers are much more common.

To our knowledge, there are few CBIR systems in dermatology. Chung et al. [12] created a skin cancer database. Users can query the database by feature attribute values (shape and texture), or by synthesised image colours. It does not include a query-by-example method, as do most common CBIR systems. The report concentrates on the description of the web-based browsing and data

mining. However, nothing is said about database details (number, lesion types, acquisition technique), nor about the performance of the retrieval system. Celebi et al. [13] developed a system for retrieving skin lesion images based on shape similarity. The novelty of that system is the incorporation of human perception models in the similarity function. Results on 184 skin lesion images show significant agreement between computer assessment and human perception. However, they only focus on silhouette shape similarity and do not include many features (colour and texture) described in other papers by the same authors [11]. Rahman et al. [14] presented a CBIR system for dermatoscopic images. Their approach include image processing, segmentation, feature extraction (colour and textures) and similarity matching. Experiments on 358 images of pigmented skin lesions from three categories (benign, dysplastic nevi and melanoma) are performed. A quantitative evaluation based on the precision curve shows the effectiveness of their system to retrieve visually similar lesions (average precision  $\simeq 60\%$ ). Dorileo et al. [15] presented a CBIR system for wound images (necrotic tissue, fibrin, granulation and mixed tissue). Features based on histogram and multispectral co-occurrence matrices are used to retrieve similar images. The performance is evaluated based on measurements of precision ( $\simeq 50\%$ ) on a database of 215 images. All these approaches only consider a few classes of lesions and/or do not exploit many useful features in this context.

Dermatology atlases containing a large number of images are available on line [16, 17]. However, their searching tool only allows query by the name of the lesion. On the other hand, the possibility of retrieving images based on visual similarity would greatly benefit both the non-expert users and the dermatologists. As already pointed out [4, 14], there is a need for CBIR as a decision support tool for dermatologists in the form of a display of relevant past cases, along with proven pathology and other suitable information. CBIR could be used to present cases that are not only similar in diagnosis, but also similar in appearance and cases with visual similarity but different diagnoses. Hence, it would be useful as a training tool for medical students and researchers to browse and search large collection of disease related illustrations using their visual attributes.

Motivated by this, we propose a CBIR approach for skin lesion images. The present work focuses on 5 common classes of skin lesions: Actinic Keratosis (AK), Basal Cell Carcinoma (BCC), Melanocytic Nevus / Mole (ML), Squamous Cell Carcinoma (SCC), Seborrhoeic Keratosis (SK). As far we know, **this is the first query-by-example CBIR system for these 5 classes of lesions**. Our system mainly relies on colour and texture features, and gives values of precision between 59% and 63%.

## 2 Feature extraction

CBIR undertakes the extraction of several features from each image, which, consequently, are used for computing similarity between images during the retrieval procedure. These features describe the content of the image and that is why

they must be appropriately selected according to the context. The features have to be discriminative and sufficient for the description of different pathologies. Basically, the key to attain a successful retrieval system is to choose the right features that represent each class of images as uniquely as possible.

Many feature selection and extraction strategies have been proposed [6, 7] from the perspective of classification of images as malignant or benign. Different features attempt to reflect the parameters used in medical diagnosis, such as *ABCD* rule for melanoma detection [18]. These features are certainly effective for the classification purpose, as seen from the performance of some classification-based systems in this domain, claiming a correct classification up to 100% [10] or specificity/sensitivity of 92.34%/93.33% [11]. However, features good for classification or distinguishing one disease from another may not be suitable for retrieval and display of similar appearing lesions. In this retrieval system, we are looking for similar images in term of colour, texture, shape, etc. By selecting and extracting good representative features, we may be able to identify images similar to an unknown query image, whether it belongs to the same disease group or not. Similar images belonging to different classes may give an idea about the certainty of classification.

Skin lesions appear mainly characterised by their colour and texture. In the rest of this section we will describe the features that can describe such properties, as well as some normalisation procedures employed to make sure only the lesion colour information is taken into consideration.

**Colour** Colour features are represented by the mean colour  $\mu = (\mu_R, \mu_G, \mu_B)$  of the lesion and their covariance matrix  $\Sigma$ . Let

$$C_{XY} = \frac{1}{N} \left[ \sum_{i=1}^N X_i Y_i \right] - \mu_X \mu_Y \quad (1)$$

where:  $N$  is the number of pixels in the lesion,  $X_i$  the colour component of channel  $X$  ( $X, Y \in \{R, G, B\}$ ) of pixel  $i$ . Assuming to use the original *RGB* (Red, Green, Blue) colour space, the covariance matrix is:

$$\Sigma = \begin{bmatrix} C_{RR} & C_{RG} & C_{RB} \\ C_{GR} & C_{GG} & C_{GB} \\ C_{BR} & C_{BG} & C_{BB} \end{bmatrix} \quad (2)$$

In this work, *RGB*, *HSV* (Hue, Saturation, Value) and *CIE\_Lab*, *CIE\_Lch* (Munsell colour coordinate system [14]) and Otha [19] colour spaces are used.

A number of normalisation techniques have been applied before extracting colour features. We normalised each colour component by the average of the same component of the safe skin of the same patient, because it had best performance.

After experimenting with the 5 different colour spaces, we selected the normalised *RGB*, because it was slightly better, but there was not a huge difference.

**Texture** Texture features are extracted from generalised co-occurrence matrices. Assume an image  $I$  having  $N_x$  columns,  $N_y$  rows and  $N_g$  grey levels. Let  $L_x = \{1, 2, \dots, N_x\}$  be the columns,  $L_y = \{1, 2, \dots, N_y\}$  be the rows, and  $G_x = \{0, 1, \dots, N_g - 1\}$  be the set of quantised grey levels. The co-occurrence matrix  $P_\delta$  is a matrix of dimension  $N_g \times N_g$ , where [20]:

$$P_\delta(i, j) = \#\{(k, l), (m, n) \in (L_y \times L_x) \times (L_y \times L_x) | I(k, l) = i, I(m, n) = j\} \quad (3)$$

i.e. the number of co-occurrences of the pair of grey level  $i$  and  $j$  which are a distance  $\delta = (d, \theta)$  apart. In our work, the pixel pairs  $(k, l)$  and  $(m, n)$  have distance  $d = 1, \dots, 6$  and orientation  $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$ .

Generalised co-occurrence matrices (GCM) are the extension of the co-occurrence matrix to multispectral images, i.e. images coded on  $n$  colour channels. Let  $u$  and  $v$  be two colour channels. The generalised co-occurrence matrices are:

$$P_\delta^{(u,v)}(i, j) = \#\{(k, l), (m, n) \in (L_y \times L_x) \times (L_y \times L_x) | I_u(k, l) = i, I_v(m, n) = j\} \quad (4)$$

For example, in case of colour image, coded on three channels ( $RGB$ ), we have six cooccurrence matrices: (RR),(GG),(BB) that are the same as grey level co-occurrence matrices computed on one channel and (RG), (RB), (GB) that take into account the correlations between the channels.

In order to have orientation invariance for our set of GCMs, we averaged the matrices with respect to  $\theta$ . Quantisation levels  $N_G = 64, 128, 256$  are used for the three colour spaces:  $RGB$ ,  $HSV$  and  $CIE\_Lab$ .

From each GCM we extracted 12 texture features: energy, contrast, correlation, entropy, homogeneity, inverse difference moment, cluster shade, cluster prominence, max probability, autocorrelation, dissimilarity and variance as defined in [20], for a total of 3888 texture features (12 features  $\times$  6 inter-pixel distances  $\times$  6 colour pairs  $\times$  3 colour spaces  $\times$  3 grey level quantisations).

Texture features are also extracted from the sum- and difference-histograms (SDHs) as proposed by Unser [21]. These histograms are defined as:

$$h_S(i) = \#\{(k, l), (m, n) \in (L_y \times L_x) \times (L_y \times L_x) | I(k, l) + I(m, n) = i\} \quad (5)$$

$$h_D(j) = \#\{(k, l), (m, n) \in (L_y \times L_x) \times (L_y \times L_x) | I(k, l) - I(m, n) = j\} \quad (6)$$

As for the co-occurrence matrices, we generalised the SDHs by considering the intra- and inter-plane sum- and difference-histograms:

$$h_{S,D}^{(u,v)}(i) = \#\{(k, l), (m, n) \in (L_y \times L_x) \times (L_y \times L_x) | I_u(k, l) \pm I_v(m, n) = i\} \quad (7)$$

Similarly to the GCMs, we constructed a set of SDHs varying pixel displacement, orientation, quantisation level, and colour spaces.

From each SDH we extracted 15 features: sum mean, sum variance, sum energy, sum entropy, diff mean, diff variance, diff energy, diff entropy, cluster shade, cluster prominence, contrast, homogeneity, correlation, angular second moment, entropy as defined in [21], as well as the relative illumination invariant features described by Münzenmayer [22], for a total of other 9720 features (15 features  $\times$  2 illumination invariants  $\times$  6 inter-pixel distances  $\times$  6 colour pairs  $\times$  3 colour spaces  $\times$  3 grey level quantisations).

### 3 Similarity matching

The retrieval system is based on a similarity measure defined between the query image  $Q$  and a database image  $I$ .

For colour covariance-based features, the Bhattacharyya distance metric is used as follow:

$$D_C(Q, I) = \frac{1}{8}(\mu_Q - \mu_I)^T \left[ \frac{(\Sigma_Q + \Sigma_I)}{2} \right]^{-1} (\mu_Q - \mu_I) + \frac{1}{2} \ln \frac{|\frac{(\Sigma_Q + \Sigma_I)}{2}|}{\sqrt{|\Sigma_Q||\Sigma_I|}} \quad (8)$$

where  $\mu_Q$  and  $\mu_I$  are the average colour feature vectors,  $\Sigma_Q$  and  $\Sigma_I$  are the covariance matrices of the lesion of  $Q$  and  $I$  respectively, and  $|\cdot|$  denotes the matrix determinant.

The Euclidean distance  $D_T(Q, I)$  is used for distances between a subset of texture features  $f_{subset}$ , selected as described later.

$$D_T(Q, I) = \|f_{subset}^Q - f_{subset}^I\| = \sqrt{\sum_{i=1}^m (f_i^Q - f_i^I)^2} \quad (9)$$

Other metric distances have been considered, but gave worse results.

We aggregated the two distances into a similarity matching function as:

$$S(Q, I) = w_C \cdot D_C(Q, I) + (1 - w_C) \cdot D_T(Q, I) \quad (10)$$

where  $w_C$  and  $w_T$  are weighting factors that need to be selected experimentally. In our case,  $w_C = 0.7$  gave the best results.

#### 3.1 Feature selection

Feature selection is applied in order to select a subset of texture features from all the feature extracted. Two feature selection algorithms have been implemented: a greedy algorithm and a genetic algorithm (GA) [23]. In both cases, the objective function is the maximisation of the number of correctly retrieved images, i.e. the images belonging to the same class as the query image. This measure is closely related to *precision*, that is the ratio of the number of relevant images returned to the total number of images returned. We averaged it using each image in the database as query image, and asking the system to retrieve 10 similar images for each presented image (not retrieving itself). Often, in the information retrieval context, the F-measure, that is a weighted harmonic mean of precision and recall is often used. In our case, due to the uneven distribution of lesions into classes, it seemed more appropriate to maximise the precision.

The greedy algorithm was stopped once 10 features were selected. The greedy algorithm was slightly modified so that features with high correlation with any of the already selected features were eliminated.

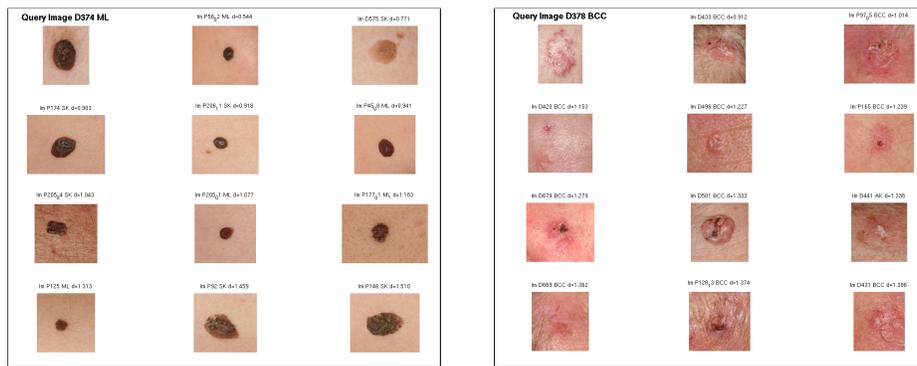
In the genetic algorithm, the feature indexes are encoded in the chromosomes as integer numbers. Each chromosome contains 10 features. Other GA

parameters (determined after a number of experiments varying such parameters) are: 1000 individuals, 0.9 crossover rate, adaptive feasible mutation, 100 generations, 10 runs. The results reported below are based on the GA as it had higher performance.

## 4 Results and evaluation

The effectiveness of the proposed retrieval system is evaluated on our image database of 208 lesions, belonging to 5 classes (11 AK, 59 BCC, 58 ML, 18 SCC, 62 SK). Images are acquired using a Canon EOS 350D SRL camera, having a resolution of about 0.03 mm. Lesions are manually segmented in this study. Ground truth is provided by the medical co-authors.

Typical screen-shots of our CBIR system are shown in Figure 1.

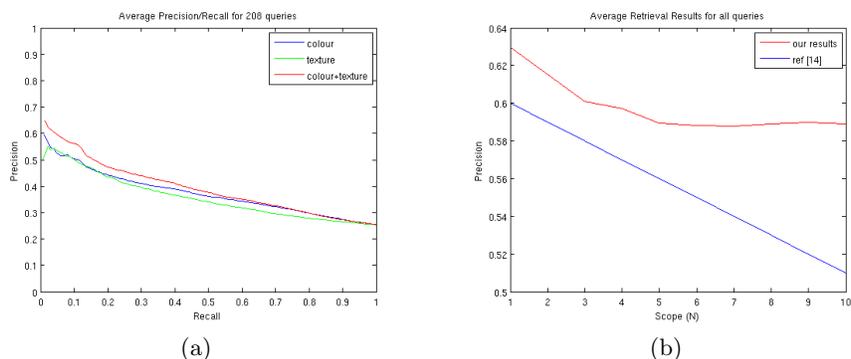


**Fig. 1.** Two screenshots showing retrieved images similar to the query image (top left image in each screenshot). Respectively 5/11 and 10/11 are correctly retrieved.

For medical image retrieval systems, the evaluation issue is very often neglected in most of the papers [4]. We show average precision/recall and precision/scope curves obtained by evaluating top  $N$  retrieved results (scope). These graphs justify our choices of best parameters and techniques for our system.

Figure 2(a) shows the precision/recall curves obtained using colour and texture features, where a subset of the latter ones are selected by GA (which outperforms the greedy feature selection). Note that colour and texture feature are co-optimised so there may be some better independent colour and texture feature sets.

Figure 2(b) shows the precision/scope curves using our method, and the average curve obtained by a similar retrieval system [14]. We estimated the average curve assuming even lesion distribution between the three classes they consider. Our system precision is about 8% higher of their system one. However a fair comparison is not possible, because their lesion classes are completely different from ours, only one of which (mole) is one of the classes we consider.



**Fig. 2.** (a) Precision/Recall curves using colour and texture features, (b) Precision/Scope curves using our method and the one reported in ref. [14]

## 5 Conclusions

We have presented a CBIR system as a diagnostic aid for skin lesion images. We believe that presenting images with known pathology that are visually similar to an image being evaluated may provide intuitive clinical decision support to dermatologists. Further studies will include the extraction of other texture-related features (i.e. fractal dimension, Gabor- and Tamura-based) as well as shape and boundary features. A larger database is being reached. We plan also to include relevance feedback, which is commonly used in image retrieval, but has not yet been used for medical images.

**Acknowledgements:** We thank the Wellcome Trust for funding this project.

## References

1. Rui, Y., Huang, T.S., Chang, S.F.: Image retrieval: Current techniques, prominent directions, and open issues. *Journal of Visual Communication and Image Representation* **10** (1999) 39–62
2. Smeulders, A.W.M., Member, S., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(12) (2000) 1349–1380
3. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys* **40**(2) (April 2008) 5:1–5:60
4. Müller, H., Michoux, N., Bandon, D., Geissbühler, A.: A review of content-based image retrieval systems in medical applications - clinical benefits and future directions. *International Journal of Medical Informatics* **73** (2004) 1–23
5. Celebi, M.E., Iyatomi, H., Schaefer, G., Stoecker, W.V.: Lesion border detection in dermoscopy images. *Computerized Medical Imaging and Graphics* **33**(2) (2009) 148–153
6. Wollina, U., Burroni, M., Torricelli, R., Gilardi, S., Dell’Eva, G., Helm, C., Bardey, W.: Digital dermoscopy in clinical practise: a three-centre analysis. *Skin Research and Technology* **13** (May 2007) 133–142(10)

7. Seidenari, S., Pellacani, G., Pepe, P.: Digital videomicroscopy improves diagnostic accuracy for melanoma. *Journal of the American Academy of Dermatology* **39**(2) (1998) 175–181
8. Lee, T.K., Claridge, E.: Predictive power of irregular border shapes for malignant melanomas. *Skin Research and Technology* **11**(1) (2005) 1–8
9. Schmid-Saugeons, P., Guilloid, J., Thiran, J.P.: Towards a computer-aided diagnosis system for pigmented skin lesions. *Computerized Medical Imaging and Graphics* **27** (2003) 65–78
10. Maglogiannis, I., Pavlopoulos, S., Koutsouris, D.: An integrated computer supported acquisition, handling, and characterization system for pigmented skin lesions in dermatological images. *IEEE Transactions on Information Technology in Biomedicine* **9**(1) (2005) 86–98
11. Celebi, M.E., Kingravi, H.A., Uddin, B., Iyatomi, H., Aslandogan, Y.A., Stoecker, W.V., Moss, R.H.: A methodological approach to the classification of dermoscopy images. *Computerized Medical Imaging and Graphics* **31**(6) (2007) 362 – 373
12. Chung, S.M., Wang, Q.: Content-based retrieval and data mining of a skin cancer image database. In: *International Conference on Information Technology: Coding and Computing (ITCC'01)*, Los Alamitos, CA, USA, IEEE Computer Society (2001) 611–615
13. Celebi, M.E., Aslandogan, Y.A.: Content-based image retrieval incorporating models of human perception. *Information Technology: Coding and Computing, International Conference on* **2** (2004) 241
14. Rahman, M.M., Desai, B.C., Bhattacharya, P.: Image retrieval-based decision support system for dermatoscopic images. In: *IEEE Symposium on Computer-Based Medical Systems*, Los Alamitos, CA, USA, IEEE Computer Society (2006) 285–290
15. Dorileo, E.A.G., Frade, M.A.C., Roselino, A.M.F., Rangayyan, R.M., Azevedo-Marques, P.M.: Color image processing and content-based image retrieval techniques for the analysis of dermatological lesions. *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS 2008)* (August 2008) 1230–1233
16. : Dermnet: the dermatologist's image resource (2007) *Dermatology Image Atlas*, available at: <http://www.dermnet.com/>.
17. Cohen, B.A., Lehmann, C.U.: *Dermatlas* (2000-2009) *Dermatology Image Atlas*, available at: <http://dermatlas.med.jhmi.edu/derm/>.
18. Johr, R.H.: Dermoscopy: alternative melanocytic algorithms—the abcd rule of dermoscopy, menzies scoring method, and 7-point checklist. *Clinics in Dermatology* **20**(3) (May-June 2002) 240–247
19. Ohta, Y.I., Kanade, T., Sakai, T.: Color information for region segmentation. *Computer Graphics and Image Processing* **13**(1) (July 1980) 222 – 241
20. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics* **3**(6) (1973) 610–621
21. Unser, M.: Sum and difference histograms for texture classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **8**(1) (January 1986) 118–125
22. Munzenmayer, C., Wilharm, S., Hornegger, J., Wittenberg, T.: Illumination invariant color texture analysis based on sum- and difference-histograms. In: *27th, Vienna, Austria, August 31 - September 2, 2005. Springer BibRef LNCS3663 German Pattern Recognition Symposium.* (2005) 17–24
23. Goldberg, D.E.: *Genetic Algorithms in Search, Optimization, and Machine Learning.* Addison-Wesley, Reading, MA (1989)