Content-Based Image Retrieval of Skin Lesions by Evolutionary Feature Synthesis

Content-Based Image Retrieval of Skin Lesions by Evolutionary Feature Synthesis

Lucia Ballerini[1], Xiang Li[1], Robert B. Fisher[1], Ben Aldridge[2], and Jonathan Rees[2]

[1] School of Informatics, University of Edinburgh, UK
x.li-29@sms.ed.ac.uk, lucia.ballerini@ed.ac.uk, rbf@inf.ed.ac.uk
[2] Dermatology, University of Edinburgh, UK
ben.aldridge@ed.ac.uk, jonathan.rees@ed.ac.uk

Abstract. This paper gives an example of evolved features that improve image retrieval performance. A content-based image retrieval system for skin lesion images is presented. 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. Evolutionary algorithms are used to create composite features that optimise a similarity matching function. Experiments on our database of 533 images are performed and results are compared to those obtained using simple features. The use of the evolved composite features improves the precision by about 7%.

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. A query-by-example CBIR involves providing the CBIR system with an example image and retrieves the most visually similar images. This is our goal as described later.

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. Their 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 online [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. 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 [4, 14]. 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). Our system mainly relies on colour and composite texture features, evolved using genetic algorithms, and gives values of precision between 67% and 82%. The use of the evolved composite features improves the precision by about 7%. The structure of the paper is as follows. Section 2 defines the simple features. Section 3 is devoted our new proposal. Section 4 defines the similarity criteria. Results are presented in 5. Conclusions follow.

2 Feature extraction

CBIR requires 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 attaining a successful retrieval system is to choose the right features that represent each class of images as uniquely as possible. Many feature 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 the *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 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. Skin lesions appear mainly characterised by their colour and texture. In this section we will describe simple features that can capture such properties. Later we will describe how to evolve composite features from these simple ones.

Colour features are represented by the mean colour $\mu = (\mu_R, \mu_G, \mu_B)$ of the lesion and their covariance matrix Σ. Let $\mu_X = \frac{1}{N}\sum_{i=1}^{N} X_i$ and $C_{XY} = \frac{1}{N}\left[\sum_{i=1}^{N} X_i Y_i\right] - \mu_X \mu_Y$, 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: $\Sigma = \begin{bmatrix} C_{RR} & C_{RG} & C_{RB} \\ C_{GR} & C_{GG} & C_{GB} \\ C_{BR} & C_{BG} & C_{BB} \end{bmatrix}$. 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 healthy skin of the same patient, because it had best performance. After experimenting with the 5 different colour spaces, we choose the normalised RGB, because it gave slightly better results than the other colour spaces.

Texture features are extracted from generalised co-occurrence matrices (CGM), that are the extension of the co-occurrence matrix [20] to multispectral images. Assume an image I having N_x columns, N_y rows and N_g grey levels. Let $L_x = \{1, 2, \cdots, N_x\}$ be the columns, $L_y = \{1, 2, \cdots, N_y\}$ be the rows, and $G_x = \{0, 1, \cdots, N_g - 1\}$ be the set of quantised grey levels. Let u and v be two colour channels. The generalised co-occurrence matrices are: $P_\delta^{(u,v)}(i,j) = \#\{((k,l),(m,n)) \in (L_y \times L_x) \times (L_y \times L_x) | I_u(k,l) = i, I_v(m,n) = j\}$ 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 $d = 1, \cdots, 6$ and orientation $\theta = 0°, 45°, 90°, 135°$, i.e. $(m = k{+}d, n = l)$, $(m = k{+}d, n = l{+}d)$, $(m = k, n = l{+}d)$, $(m = k{-}d, n = l{+}d)$. In order to have orientation invariance for our set of GCMs, we averaged the matrices with respect to θ. 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]. We generalised the SDHs by considering the intra- and inter-plane sum- and difference-histograms: $h_{S,D}^{(u,v)}(i) = \#\{((k,l),(m,n)) \in (L_y \times L_x) \times (L_y \times L_x) | I_u(k,l) \pm I_v(m,n) = i\}$. 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 Evolutionary Feature Synthesis

Evolutionary algorithms have already been applied to feature synthesis problems. Aurnhammer [23] and Lam et al. [24] described the use of genetic programming to generate texture features and reported very promising results on image classification problems. Li et al. [25] proposed a hybrid of a co-evolutionary genetic programming and expectation maximisation algorithm applied on partially labelled data. They show that their algorithm outperforms support vector machines in the sense of both the classification performance and the computational efficiency in the testing phase.

In our work, each synthesised feature is derived by combining simple features using a series of operators. A genetic algorithm (GA) [26] is used in this phase.

The main issues in applying a GA to any problem are selecting an appropriate encoding representation of the solutions, defining an adequate evaluation function (fitness), and choosing the values of the parameters used by the algorithm (e.g. population size, crossover, etc.). In the case of synthesised features there are two basic items: the index of the simple features (among the 13608 extracted) to be selected and the operators used to combine them. Each chromosome is composed of two parts: a part which encodes the index set of the simple features and a part which encodes the operators. In this work we present results obtained using 6 operators: $\{1, 2, +, -, *, /\}$. Each operator is applied to a pair of features. The first 2 operators mean that only the first or the second features of the pair is chosen. The last 4 operators perform the given mathematical operation on the two features of the pair. The fitness is the number of correctly retrieved images, i.e. the images belonging to the same class as the query image. 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).

In the GA, the feature indexes and the operators are encoded in the chromosomes as integer numbers. Each chromosome contains 10 features and 5 operators (one for each pair of the 10 features). The implementation of mutation and crossover on integer numbers is straightforward, with the condition to generate children satisfying the range

and integer constraints on decision variables. Other GA parameters (determined after a number of experiments varying such parameters) are: 200 individuals, 0.9 crossover rate, 0.01 mutation rate, stochastic uniform selection. The stopping criteria is upon reaching the maximum number of generations (30) or having a change in the fitness of less than 10^{-6}. Results reported later are the average over 20 runs.

4 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 $D_C(Q,I) = \frac{1}{8}(\mu_Q - \mu_I)^T \left[\frac{(\Sigma_Q+\Sigma_I)}{2}\right]^{-1}(\mu_Q - \mu_I) + \frac{1}{2}\ln\frac{\left|\frac{(\Sigma_Q+\Sigma_I)}{2}\right|}{\sqrt{|\Sigma_Q||\Sigma_I|}}$ is used, where μ_Q and μ_I are the average colour (over all pixels in the lesion) feature vectors, Σ_Q and Σ_I are the covariance matrices of the lesion of Q and I respectively (computed as described in Section 2), and $|\cdot|$ denotes the matrix determinant. The Euclidean distance $D_T(Q,I) = \|f^Q_{comp} - f^I_{comp}\| = \sqrt{\sum_{i=1}^{m}(f^Q_i - f^I_i)^2}$ is used for distances between the composite features f_{comp}, evolved as previously described. where m is the number of features: $m = 5$ for the composite features, $m = 10$ for the simple features used for comparison.

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

$$S(Q,I) = w_C \cdot D_C(Q,I) + (1 - w_C) \cdot D_T(Q,I) \tag{1}$$

where w_C is a weighting factor that has been selected experimentally, after trying all the values: $\{0.1, 0.2, \cdots, 0.9\}$. In our case, $w_C = 0.7$ gave the best results.

5 Results and evaluation

Our image database comprises 533 lesions, belonging to 5 classes (20 AK, 116 BCC, 224 ML, 20 SCC, 153 SK). Images are acquired using a Canon EOS 350D SRL camera, having a resolution of about 0.03 mm. Lesions are segmented using the method described in [27]. The ground truth used for the experiments is based on agreed classifications by 2 dermatologists. Feature synthesis is performed using only 100 images (20 for each class randomly chosen). The effectiveness of the proposed retrieval system is then evaluated on the entire database.

One example of the composite feature set is shown in Figure 1, together with the plot of class distribution of one of them, where it can be seen it slightly distinguishes ML and SK from AK, BCC, SCC. A typical screen-shot of our CBIR system is shown in Figure 2(a).

For medical image retrieval systems, the evaluation issue is very often neglected in most of the papers [4]. In an information retrieval scenario, *precision* is defined as the number of relevant documents retrieved by a search divided by the total number of documents retrieved by that search (*scope*), and *recall* is defined as the number of relevant documents retrieved by a search divided by the total number of existing relevant documents. We show average precision/scope curves obtained by evaluating top N retrieved

#	colours	feature name	dist	q.level	operator
1	BB	homogeneity ii	4	64	+
	HS	entropy	3	64	
2	ab	correlation	2	128	-
	HH	diff entropy ii	5	256	
3	aa	diff variance ii	2	64	-
	bb	entropy ii	5	128	
4	RB	inv diff moment	3	128	1
5	VV	diff energy ii	4	128	+
	La	cluster prom	3	256	

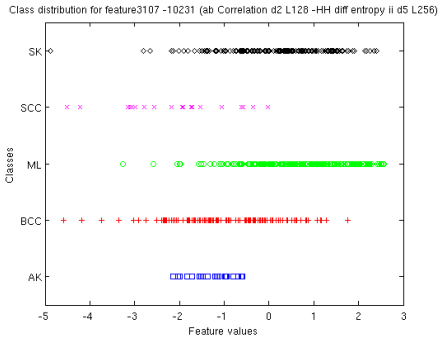


Fig. 1. Example of evolved composite features, and class distribution of feature 2

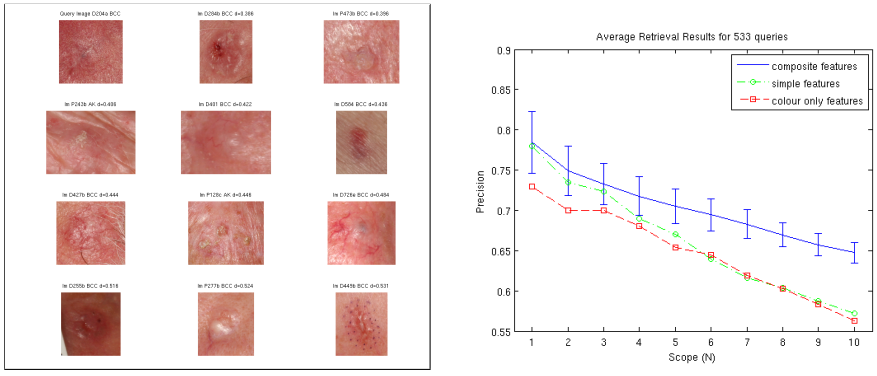
results (scope). We compare our results with the results obtained by the same system except using simple features. The simple features are selected by a GA (using the same parameters as the other GA). Similarity function (1) is used with $m = 10$ in D_T. Our 5 composite features originate from 10 simple features, therefore we decided to compare 5 composite features against 10 simple ones.

Figure 2(b) shows the precision/scope curves obtained using the composite features synthesised by our method. Precision/scope curves obtained using simple features and only colour features are shown for comparison. Note that using $m = 5$ composite features outperform $m = 10$ simple features and that the use of the composite features improves the precision by about 7%, where at scope=1 the difference is 1%. The comparison with the performances obtained using only the colour features makes clear the improvement of our system due to the composite texture features.

As far we know, our system is the first query-by-example CBIR system for these 5 classes of lesions, therefore comparison with other system is not possible. The use of a system developed for generic image retrieval gave very poor results on our data.

6 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. We have shown that the use of evolved composite features improves the performance of the system compared to the use of a larger number of standard features. Given the encouraging results obtained using a small set of feature combination operators we plan to investigate the use of a larger number of operators that combine an arbitrary number of features. Genetic programming (GP) may offer several advantages over GA. Further studies will also include the extraction of other texture-related features (i.e. fractal dimension, Gabor- and Tamura-based) as well as shape and boundary features. We plan also to include relevance feedback, which is commonly used in image retrieval, but has not yet been used for medical images.



(a) (b)

Fig. 2. (a) A screenshot showing retrieved images similar to the query image (top left image). (b) Precision/Scope curves using our evolved composite features, simple features and colour only features. Vertical bars report performance over 20 runs (mean $\pm$ std).

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

References

1. Rui, Y., Huang, T.S., Chang, S.F.: Image retrieval: Current techniques, prominsign 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., Geissbuhler, 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., Guillod, 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 2001), 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. In: International Conference on Information Technology: Coding and Computing (ITCC 2004). Volume 2., Los Alamitos, CA, USA, IEEE Computer Society (2004) 241–245
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 Altas, available at: http://www.dermnet.com/.
17. Cohen, B.A., Lehmann, C.U.: Dermatlas (2000-2009) Dermatology Image Altas, available at: http://dermatlas.med.jhmi.edu/derm/.
18. Johr, R.H.: Dermoscopy: alternative melanocytic algorithms–the ABCD rule of dermatoscopy, 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., Shanmungam, 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 Kropatsch, W., Sablatnig, R., Hanbury, A., eds.: 27th German Pattern Recognition Symposium (DAGM 2005). Volume 3663 of LNCS., Vienna, Austria (2005) 17–24
23. Aurnhammer, M.: Evolving texture feature by genetic programming. In Giacobini, M.e.a., ed.: EvoWorkshops 2007. Volume 4448 of LNCS. (2007) 351–358
24. Lam, B., Ciesielski, V.: Discovery of human-competitive image texture feature extraction programs using genetic programming. In Deb, K., Poli, R., Banzhaf, W., Beyer, H.G., Burke, E., Darwen, P., Dasgupta, D., Floreano, D., Foster, J., Harman, M., Holland, O., Lanzi, P.L., Spector, L., Tettamanzi, A., Thierens, D., Tyrrell, A., eds.: Genetic and Evolutionary Computation – GECCO-2004, Part II. Volume 3103 of Lecture Notes in Computer Science., Seattle, WA, USA, Springer-Verlag (26-30 June 2004) 1114–1125
25. Li, R., Bhanu, B., Dong, A.: Feature synthesized EM algorithm for image retrieval. ACM Transaction on Multimedia Computing Communications and Applications **4**(2) (2008) 10:1–10:24
26. Goldberg, D.E.: Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, Reading, MA (1989)
27. Xiang, L., Aldridge, B., Ballerini, L., Fisher, R., Rees, J.: Depth data improves skin lesion segmentation. In G.-Z. Yang et al., ed.: Proc. 12th Int. Conf. on Medical Image Computing and Computer Assisted Intervention (MICCAI), London. (2009) 1100–1107