A Comprehensive Descriptor of Shape: Method and Application to Content-Based Retrieval of Similar Appearing Lesions in Medical Images

Jiajing Xu · Jessica Faruque · Christopher F. Beaulieu · Daniel Rubin · Sandy Napel

Published online: 6 May 2011
© Society for Imaging Informatics in Medicine 2011

Abstract We have developed a method to quantify the shape of liver lesions in CT images and to evaluate its performance for retrieval of images with similarly-shaped lesions. We employed a machine learning method to combine several shape descriptors and defined similarity measures for a pair of shapes as a weighted combination of distances calculated based on each feature. We created a dataset of 144 simulated shapes and established several reference standards for similarity and computed the optimal weights so that the retrieval result agrees best with the reference standard. Then we evaluated our method on a clinical database consisting of 79 portal-venous-phase CT liver images, where we derived a reference standard of similarity from radiologists' visual evaluation. Normalized Discounted Cumulative Gain (NDCG) was calculated to compare this ordering with the expected ordering based on the reference standard. For the simulated lesions, the mean NDCG values ranged from 91% to 100%, indicating that our methods for combining features were very accurate in representing true similarity. For the clinical images, the mean NDCG values were still around 90%, suggesting a strong correlation between the computed similarity and the independent similarity reference derived the radiologists.

Keywords Image retrieval · Image analysis · Image processing

Introduction

Due in part to the increasing number of images per patient study, it is ever more challenging for diagnostic radiologists to maintain accuracy and interpretation efficiency when reading cross-sectional imaging exams [1, 2]. We believe it would be possible to improve the diagnostic process by enabling radiologists to compare the images being interpreted with prior cases whose diagnoses have been established as the basis for decision support. However, to establish this capability requires tools and databases that would allow rapid access to prior cases that are similar in terms of shared imaging features to the images they are interpreting. *Visual similarity* is a subjective combination of several observable features such as shape, boundary sharpness, intensity, and texture. Lesion shape is a particularly important feature that is used to determine lesion similarity. In this paper, we concentrate on the shape of the borders of lesions and present a computable generalized shape descriptor and evaluate its use in the retrieval of CT images of liver lesions.

Previous work [3] on shape analysis of liver lesions focused on measuring convex hull depth, irregularity, and jag count of the lesion, and found weak correlations between radiologists' opinion and these shape descriptors individually. However, this work limited the use of shape descriptors for a binary classification of lesion type (e.g., benign or malignant) and did not evaluate them for use in similar image retrieval, wherein a system would return a ranked series of images from most to least similar to a query image.

Many different shape descriptors have been defined [4–7] but each is suitable only for particular kinds of shapes. In this paper, we introduce a *generalized shape descriptor*, which combines several specific descriptors with arbitrary weights

J. Xu · J. Faruque
Department of Electrical Engineering, Stanford University,
Stanford, CA 94305, USA

C. F. Beaulieu · D. Rubin · S. Napel
Department of Radiology, Stanford University,
Stanford, CA 94305, USA

J. Xu (✉)
119 Quillen Court Apt. 202,
Stanford, CA 94305, USA
e-mail: jiajing@stanford.edu

that can be trained to rank a set of images based on a reference standard of visual similarity. Our method is based on the least absolute shrinkage and selection operator (LASSO) [8], which combines multiple descriptors to obtain a new compact descriptor. The contribution of each descriptor to this final descriptor is determined according to their relative performance, resulting in the most relevant descriptors having the greatest influence. In the following sections, we define the individual components of the shape descriptors, describe the use of a technique for training their weights, and describe and report on experiments with image data, both simulated and from a database of liver CT scans.

Materials and Methods

Shape Features

Radial Distance Signature

Once we obtained the coordinates $x(n)$ and $y(n)$ of the pixels that make the shape boundary (see Section “[Feature Extraction](#)”), we derive a one-dimensional shape signature $r(n)$, expressed by the distance of the boundary points from the centroid (x_c, y_c) of the shape.

$$x_c = \frac{1}{N} \sum_{n=0}^{N-1} x(n), y_c = \frac{1}{N} \sum_{n=0}^{N-1} y(n) \quad (1)$$

$$r(n) = \sqrt{(x(n) - x_c)^2 + (y(n) - y_c)^2} \quad (2)$$

The signature is then normalized by the maximum distance between the centroid and the n boundary points, which makes the signature invariant to rotation and scale.

$$r^{\text{NORM}}(n) = \frac{r(n)}{\max_i r(i)} \quad (3)$$

Two Simple Shape Descriptors

We used two common global shape descriptors: compactness and boundary roughness. Compactness [9] is defined as $C=A/P^2$, where A is the area and P is the perimeter of the shape; the average boundary roughness [10] was calculated by dividing the radial distance signals into small segments of equal length and then estimating a roughness index for each one of them according to the formula.

$$R(j) = \sum_{i=j}^{L+j} |r(i) - r(i+1)|, j = 1, \dots, \left\lfloor \frac{N}{L} \right\rfloor \quad (4)$$

$$R_{\text{avg}} = \left\lfloor \frac{L}{N} \right\rfloor \sum_{j=1}^{\left\lfloor \frac{L}{N} \right\rfloor} R(j) \quad (5)$$

where $r(i)$ is the radial distance signal, $R(j)$ is the roughness index for the j th segment, L the number of boundary points in each segment, k denotes the last segment of the length L , and N is the total number of boundary points available. The roughness measure R_{avg} is calculated by averaging the roughness index over the entire lesion boundary.

Local Area Integral Invariant Signature

Other shape signatures that are rotation and scale invariant include curvature and the local area integral invariant (LAII) [11, 12]. Curvature is calculated based on differential operations and is sensitive to noise because the calculation of the curvature is dependent on second-order derivatives. In contrast to curvature, integral invariants, such as area integral invariant, employ integral operations for their calculation and are less sensitive to noise. Integral invariants are quite general and contain regularized differential invariants as a subset.

Area integral invariant: define $K_r: \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \{0,1\}$ as an indicator kernel function on the interior of a circle with radius r centered at point p :

$$K_r(p, x) = \begin{cases} 1 & |p - x| \leq r \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

For any given radius r , the corresponding integral invariant,

$$I_r(p) = \int_B K_r(p, x) dx \quad (7)$$

This can be thought of as a function from interval $[0, L]$ to $\mathbb{R}^+$ (since area is always nonnegative), bounded above by the area of B , where L is the arc length of S and the binary mask B indicates the interior of the region bounded by S . The area integral invariant can also be normalized by the area of $K_r(p)$ for convenience:

$$I_r(p) = \frac{\int_B K_r(p, x) dx}{\int_{\mathbb{R}^2} K_r(p, x) dx} \quad (8)$$

The normalized integral invariant is then bounded between 0 and 1. This is illustrated in Fig. 1.

As defined in $I_r(p)$, a scale is associated with LAII. Varying r from zero to a maximum radius r_{max} so that the local kernel $K_r(p)$ encloses the entire curve, we can generate multiscale LAII. As shown in Fig. 2, if the scale is too small, the signature becomes much noisier, and if the scale is too large, all parts of the signature tend to be identical. Hence, the scale parameter is crucial to LAII. To

Fig. 1 Illustration of LAII signature. $K_r(p)$ and $K_r(p')$ are two indicator kernel functions on the interior of a circle with radius r centered at p and p' on the boundary of the shape B . The shaded area is where the value where the kernel functions is non-zero

ensure we cover a wide range of scales, we used the LAII signature at five scales: $r_{\max}/2$, $r_{\max}/3$, $r_{\max}/5$, $r_{\max}/8$, $r_{\max}/10$.

Combining Features and Similarity Measures

We used a total of 14 features for characterizing each shape, as listed in Table 1: compactness (length of 1), the roughness (length of 1), the mean and standard deviation of the radial distance signature (RDS) (length of 2), the mean and standard deviation of the LAII at five different scales (total length of 10). We used the LASSO [8] algorithm to learn optimal weights to linearly combine each component. LASSO is a penalized regression method that improves upon ordinary least squares and ridge regression. It contains a penalty term that automatically enforces the presence of (many) zero weights. Among such zero-enforcing penalties, the L1 norm of f is the only convex one [13], thereby providing feasible algorithms for

Fig. 2 The multiscale LAII of the lesion contour shown in Fig. 1, the radius r of kernel circle varies from $0.2R$ to $0.8R$, where R is the maximum distance between any two points on the boundary. (Dark solid) $r=0.2R$; (light solid) $r=0.5R$; (dark dashed) $r=0.7R$; (light dashed) $r=0.8R$. Too large a scale would generate a flat curve. As r approaches 0, LAII becomes curvature, which will be much noisier

Table 1 List of features extracted and used

Feature ID	Feature name
1	LAII, mean, $r=R/10$
2	LAII, SD, $r=R/10$
3	LAII, mean, $r=R/8$
4	LAII, SD, $r=R/8$
5	LAII, mean, $r=R/5$
6	LAII, SD, $r=R/5$
7	LAII, mean, $r=R/3$
8	LAII, SD, $r=R/3$
9	LAII, mean, $r=R/2$
10	LAII, SD, $r=R/2$
11	RDS, mean
12	RDS, SD
13	Compactness
14	Roughness

high-dimensional data. The loss function of LASSO regression is defined as:

$$L = \sum_i \left(y_i - \sum_p \beta_p x_{ip} \right)^2 + \lambda \sum_p \|\beta_p\|_1 \quad (9)$$

where x_{ip} denotes the p th feature descriptor in the i th training case, denotes the reference standard for i th training case, and β_p denotes the weight coefficient of the p th feature descriptor. The L1 norm regularizer $\sum_p \|\beta_p\|_1$ in the LASSO regression typically leads to a sparse solution in the feature space, which means that the coefficients for the least relevant or redundant features become zero. A theoretical analysis in [14] indicates that LASSO regression is particularly effective when there are many irrelevant features and only has a comparatively small number of training examples.

Once the optimal weights are obtained, we define the similarity of a pair of shapes as the inverse of a weighted sum of differences between corresponding elements of the respective feature vectors that describe them.

Datasets

Simulated Dataset

We created a dataset of simulated shapes to learn the optimal weights for combining the shape descriptors and to use in our evaluation. Since liver lesions are generally ovoid and lobular, we generated shapes by controlling ellipse eccentricity and by modulating the ellipses with sinusoids of different amplitudes and frequencies. Manipulating these three parameters resulted in a total of 16 shapes, as shown in Fig. 3. To test our descriptors invariance to scale, we expanded this dataset to 48 shapes

Fig. 3 A set of synthetic lesion contours obtained by simulation. The complete set consisted of the basic at three scales ($\times 1$, $\times 2$, and $\times 4$) and three rotations/scale (0, 120°, 240°)

Frequency Other Parameters	1 cycle per shape	2 cycles per shape	3 cycles per shape	4 cycles per shape
Eccentricity = 1.2 Amplitude = 5% of the Major Axis
Eccentricity = 1.2 Amplitude = 10% of the Major Axis
Eccentricity = 1.6 Amplitude = 5% of the Major Axis
Eccentricity = 1.6 Amplitude = 10% of the Major Axis

by adding two scaled versions of the 16-shape base set (scale factors of 2 and 4). Similarly, to ensure a rotational invariance, we added rotated versions of the 48-shape set (rotated by 120°, 240°). The final simulated shape dataset contained $16 \times 3 \times 3 = 144$ shapes.

Clinical Dataset

Under IRB approval for retrospective analysis of de-identified patient images, we selected 79 portal venous phase CT images from our clinical archive in which liver lesions had been identified. The 79 lesions (25 cysts, 24 metastases, 14 hemangiomas, three abscesses, one fat deposition, one laceration, five focal nodular hyperplasia, and six hepatocellular carcinomas) were obtained from 15 patients (eight men, seven women), mean age 56 years old (range, 39–88 years). These types of lesions are common and span a range of image appearances. The liver CT images were acquired during the time period of February 2007 to August 2008 and used the following range of scan parameters: 120 kVp, 140–400 mAs, 2.5–5-mm slice thickness. A radiologist (C.F.B.) with 15 years of abdominal CT experience used OsiriX to circumscribe each lesion boundary.

Feature Extraction

For the simulated dataset, the coordinates of 400 lesion boundary points for each shape were obtained through the parametric equations describing it.

For the clinical dataset, the radiologist circumscribed lesions using the polygon tool in OsiriX, which stores the boundary in the format of an array of control points, typically around 10 per lesion. Since liver lesions vary significantly in size, it is necessary to apply an extra normalization step for small lesions in order to get enough points along the boundary. We first determined the size of the bounding box for the lesion, and if the shorter side of the bounding box was less than 80 pixels, we multiplied the coordinates of the control points by a factor that makes the

shorter side of the bounding box 200-pixels wide. We then fitted cubic splines through control points and sampled 40 boundary points along each spline segment that connects two adjacent control points [15]. After completely tracking along the boundary, we obtained the coordinates $x(n)$ and $y(n)$ of the boundary points. Thus, this interpolation method resulted in approximately 400 points along the boundary of the clinical lesions.

We computed the radial distance signal using Eqs. 1–3 in the section “Shape Features.” We obtained compactness by computing the area and perimeter directly from the boundary points $x(n)$ and $y(n)$ by using the MATLAB image processing toolbox. The average boundary roughness was computed using Eqs. 4 and 5. For LAII, we first constructed a small binary image containing the boundary of lesion only. We set the pixels on the interior of the lesion to be 1 and the background to be 0, followed by convolving this binary image with the kernel K_r . Evaluating the result of this convolution on each boundary point yields the LAII.

Reference Standard

Simulated Dataset

We generated six reference standards of shape similarity for the simulated shape dataset to evaluate image retrieval performance using our shape feature.

1. Reference standard 1 (RS1): For each pair of synthetic shapes, we calculated compactness for both shapes as we did in the feature extraction (“Two Simple Shape Descriptors”) and defined the similarity score as the inverse of the absolute difference of compactness for both shapes.
2. RS2: For each pair of synthetic shapes, we defined the similarity score as the inverse of the weighted sum of squared differences between the corresponding parameters that were used to generate each shape. The three parameters (lobulation amplitude, lobulation frequency, and eccentricity) are defined in the section “Simulated Dataset.” For RS2, all weights were united.

3. RS3 is similar to RS2, except the weights were 3, 1, and 1, for the eccentricity, lobulation amplitude, and lobulation frequency, respectively.
4. RS4 is similar to RS2 except the weights were 1, 3, and 1, for the eccentricity, lobulation amplitude, and lobulation frequency, respectively.
5. RS5 is similar to RS2, except the weights were 1, 1, and 3, for the eccentricity, lobulation amplitude, and lobulation frequency, respectively.
6. RS6: We fit an ellipse to each of the synthetic shapes using the fitting algorithm by Fitzgibbon [16]. Then for each pair of synthetic shapes, we defined the similarity score as the inverse of the absolute difference in the axis ratios of the two fitted ellipses.

For each reference standard, we normalized the similarity scores in the range of [1, 5], with 1 being the least similar.

Clinical Dataset

We created an independent reference standard for the 79 CT images of liver lesions described previously to enable us to evaluate image retrieval by using our trained features. Five readers (three board-certified, fellowship-trained radiologists with 20, 19, and 5 years of experience in abdominal imaging and two senior researchers in the medical imaging field) viewed each lesion image and recorded their opinions on the degree of lobulation of the lesion boundary on a scale of 1 to 5 (with 5 being very rough and lobulated and 1 being very smooth) [17]. We represented the degree of lobulation by the mean of the five scores from these readers. For each pair of lesions, we define the similarity score as the inverse of the difference of the smoothness for the two shapes. Again, the similarity scores were normalized in the range of [1, 5]. Thus, with this reference standard, a perfect retrieval system would return a sequence of images with similarities monotonically decreasing from 5 to 1.

Evaluation Measures

We used the normalized discounted cumulative gain (NDCG) [18], a method of measuring the effectiveness of information retrieval algorithms when ground truth is available, e.g., the reference standard for the simulated dataset. NDCG measures the usefulness (gain) on a scale of 0 to 1 of K retrieved lesions based on their positions in the ranked list compared to their similarity to the query lesion according to a separate reference standard. The accumulated gain is evaluated with the weight of each retrieved lesion discounted at lower ranks. Thus, for a given K , higher $NDCG(K)$ means more lesions similar to the query image are ranked ahead of dissimilar ones, with $NDCG(K)=1$ implying perfect retrieval of K images. The NDCG metric is particularly relevant in our

application since ordering in the retrieval results matters—all the most relevant retrieved images should be returned first in an ideally performing system.

We also assessed interobserver variability in the reference standard for the clinical images using Fleiss' kappa test [19]. This test evaluates the consistency of raters using categorical ratings and returns a score from 0 to 1, with 0 denoting complete disagreement and 1 denoting complete agreement.

Training and Evaluation

For both the simulated and clinical datasets, we used a leave-one-out cross-validation framework to learn weights for combining features. For example, for the simulated dataset, we withheld one of the 144 shapes and, using one of the similarity reference standards described above, with the remaining 143, we computed the optimal values for the feature weights to be used in the similarity computation using the LASSO method. These weights were then used in the similarity calculation to rank the remaining 143 lesions which, when combined with the similarity standard for the 143 lesions, generated a single NDCG curve. Finally, we computed the mean and standard deviation of NDCG over all 144 withheld shapes at each $K=1,143$. Since there is little difference among the 144 sets of feature weights, we used the average of the 144 sets of weights as the optimal feature weights corresponding to the chosen reference standard.

Results

Table 2 shows the optimal weights for the simulated dataset obtained by using the reference standards described in the section “[Reference Standard](#),” as well as the mean and the worst NDCG results over K , where K ranges from 1 to 143. Note that where we have used compactness as the reference standard, the training process results in the assignment of a single non-zero weight to the compactness feature (no. 13) as expected.

Figure 4 (left) shows the best, worst, and the mean NDCG scores for the clinical dataset using the reference standard collected from five readers (Fleiss' kappa was 0.21 (95% CI, 0.20–0.23)). Figure 5 shows a few examples for a single query image, including the top ranked seven and bottom ranked eight images, representative of the results in general; images that appear most similar to the query image were ranked higher than those that were less so, with a small number of exceptions. Perfect retrieval compared to the reference standard would produce all five ranked above a monotonically decreasing reference standard similarity (numbers in parentheses). The features that were computed

Table 2 Optimal weights learned for the simulated shapes with mean and lowest NDCG over all $K=1..143$, for various references standards

Reference standard	Optimal weights	Mean NDCG	Lowest NDCG
RS1		100%	100%
RS2		96%	94%
RS3		91%	88%
RS4		98%	95%
RS5		100%	99%
RS6		98%	97%

RS1: the inverse of the absolute difference of compactness for both shapes. RS2–RS5: the inverse of a weighted sum of squared differences between the corresponding parameters (lobulation amplitude, lobulation frequency, and eccentricity) that were used to generate each shape, where for (RS2) all weights=1. RS3: the weights were 3, 1, and 1; RS4: the weights were 1, 3, and 1; and RS5: the weights were 1, 1, and 3, for eccentricity, lobulation amplitude, and lobulation frequency, respectively. RS6: the inverse of the absolute difference in the axis ratios between ellipses fitted to each shape

to have higher weights are LAII mean at R/3, LAII standard deviation at R/2, mean of RDS, and compactness.

Discussion

Content-based image retrieval is being increasingly applied to medical images [20], with the aim of providing

radiologist interpreters with examples of known diagnoses against which to compare an unknown case. Much of this work focuses on overall image similarity, which fails to capture specific details of abnormal areas (or “lesions”) presenting within a given image or organ. Considering image features, there are potentially a large number of quantitative parameters that can be extracted from whole images or regions of interest to serve as matching features.

Fig. 4 Left: best (empty triangle), worst (dots), and average (empty circles) NDCG plotted for the dataset containing 79 lesions. For $K=5, 10$, and 15 , the combined distance achieved an average score of 89%, 90%, and 91%, respectively. Right: bar plot shows the average and standard deviation of the 79 sets of optimal weights during the leave-one-out training. There is not much variation among the 79 sets of weights

Fig. 5 Examples of retrieval for CT liver lesions (scale adjusted to fit in the figure). *Query* image is in *upper left* of each panel of 16 images, within which rankings go from highest on *upper left* to lowest on *lower right*. Due to space limitations, we only show the seven highest rankings and eight lowest rankings. Examples show similar shapes to

query image in *top half* and quite different shapes in *lower half*, as expected. Numbers above each image, $x(y)$, show x =computed similarity score and y =similarity from the reference standard, both on a scale of 1 to 5 with 5 being the best

For many diseases, the shape of the lesion and the regularity of the lesion boundary are very important in diagnosis, so we are developing methods to capture shape features reliably.

We used machine learning techniques to combine three fundamental shape descriptors and evaluated our method using both simulated lesions and actual lesions from CT data. Similarity measures derived using these methods were compared against independent reference standards for evaluation. For simulated data, we had exact mathematical representations of lesion shape against which to evaluate similarity. For clinical data, we used a reference standard of similarity derived from the radiologists' visual evaluation of lesions for similarity of lesion shapes/boundaries. For the simulated lesions, the mean NDCG values for the six shape feature variants (Table 2) ranged from 91% to 100%, indicating that our methods for combining features were very accurate in representing true similarity. For clinical lesions, the average NDCG results were lower than for simulated data, but were still around 90%, suggesting a strong correlation between the computed similarity and the independent similarity reference derived from the radiologists. We found that the calculated feature weights were strongest for four of the 14 features evaluated: LAII mean

at R/3, LAII standard deviation at R/2, mean of RDS, and compactness. The weights for these four features were quite stable across the 79-lesion dataset evaluated by the leave-one-out cross validation (Table 1; Fig. 4 (right)). This shows the utility of the LASSO method in extracting the most important components of a multielement feature vector and will enable us to focus on the strongest features in maximizing the sensitivity of the future algorithms.

This study has several limitations. Experiments in simulated datasets show that our computational shape descriptor is trainable to a wide variation of notions of similarity and, in each case, ranks similarity in close agreement with the reference standard. The results in the clinical images were visually excellent, but the mean NDCG scores were not as favorable as for the simulated data. Clinical data are much more variable in terms of image quality, and our reference standard was based on human evaluation rather than pure computation. In particular, it is challenging to arrive at a consensus reference standard for lesion shape and boundary regularity because of variations amongst viewing radiologists. Indeed, we believe that a small number of images that have high reference standard similarity to a query image but rank low in the retrieved list (e.g., Fig. 6 row 3 column 3 has

reference standard similarity with the query image of 3.5, but ranks as the sixth least similar image out of the 79 in the dataset) are caused by inaccuracies in the reference standard. We used five independent readers and averaged their results to minimize these variations; however, the usual interpretation of the Fleiss' kappa indicated only fair agreement amongst the raters. We are exploring other methods to obtain more accurate and precise reference standards for evaluating clinical datasets. Also, we did not explore variations in the shape feature vector as a function of acquisition technique; certainly, image noise and contrast, such as might be affected by scanner settings and bolus timing, will influence the computation of the shape feature vector. We only studied portal venous phase images because this is the contrast enhancement phase most commonly obtained in abdominal CT for this indication. Another limitation is that for the clinical dataset, we rely on radiologists' segmentation that consists of only a few control points, which may not be sufficient for lesions with highly lobulated boundaries or highly speculated margins. It will be desirable to have automatic boundary extraction method for more accurate delineation of the lesion margin. A final limitation is that our current method is only two-dimensional and ignores shape variations in the third dimension.

Conclusions

In conclusion, we developed a novel quantitative lesion shape descriptor that is based on machine learning techniques to combine three fundamental shape descriptors to retrieve similar-appearing lesions on CT scans of liver lesions. In our study, we evaluated performance of our descriptors using the NDCG, which accounts for ordering as well as graded truth. Our preliminary results indicate that our shape descriptor can achieve a high accuracy with both simulated and clinical datasets and compute reasonable rankings of similarity compared to a human-derived similarity standard. We believe our methods are thus promising and may effectively complement other quantitative features in building systems to retrieve similar-appearing images for decision support.

References

1. Robinson PJ: Radiology's Achilles' heel: error and variation in the interpretation of the Rontgen image. *Br J Radiol* 70:1085–98, 1997
2. Rubin GD: Data explosion: the challenge of multidetector-row CT. *Eur J Radiol* 36:74–80, 2000
3. Kim S, et al: Computer-aided image analysis of focal hepatic lesions in ultrasonography: preliminary results. *Abdominal Imaging* 34:183–191, 2009
4. Van Otterloo PJ: A contour-oriented approach to shape analysis. Prentice Hall, Englewood Cliffs, 1991
5. Zhang DS, Lu GJ: A comparative study of curvature scale space and Fourier descriptors for shape-based image retrieval. *J Vis Commun Image Represent* 14:41–60, 2003. Mar
6. Gonzalez RC, Woods RE: Digital image processing, 3rd edition. Pearson/Prentice Hall, Upper Saddle River, 2008
7. Zhang DS, Lu GJ: Review of shape representation and description techniques. *Pattern Recognition* 37:1–19, 2004
8. Tibshirani R: Regression Shrinkage and Selection Via the Lasso. *J R Stat Soc Series B Stat Methodol* 58:267–288, 1994
9. Duda RO, Hart PE: Pattern classification and scene analysis. Wiley, New York, 1973
10. Kilday J, et al: Classifying mammographic lesions using computerized image analysis. *IEEE Trans Med Imaging* 12:664–9, 1993
11. Manay S, et al: Integral invariants for shape matching. *IEEE Trans Pattern Anal Mach Intell* 28:1602–18, 2006. Oct
12. Byung-Woo H, et al: Shape representation based on integral kernels: application to image matching and segmentation in computer vision and pattern recognition. 2006 IEEE Computer Society Conference on, 2006, pp 833–840
13. Roth V: The generalized LASSO. *IEEE Trans Neural Netw* 15:16–28, 2004
14. Ng A: Feature selection, L1 vs. L2 regularization, and rotational invariance. Presented at the Proceedings of the 21st International Conference on Machine Learning. Banff, Alberta, Canada, 2004
15. Press WH: Numerical recipes in C: the art of scientific computing, 2nd edition. Cambridge University Press, Cambridge, 1992
16. Fitzgibbon AW, et al: Direct least-squares fitting of ellipses, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21(5):476–480, 1999. May
17. Napel SA, et al: Automated retrieval of CT images of liver lesions on the basis of image similarity: method and preliminary results. *Radiology* 256:243–52, 2010. Jul
18. Jarvelin K, Kekalainen J: Cumulated gain-based evaluation of IR techniques. *ACM T Inform Syst* 20:422–446, 2002
19. Fleiss JL, Cohen J: The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas* 33:613–619, 1973
20. Muller H, et al: A review of content-based image retrieval systems in medical applications-clinical benefits and future directions. *Int J Med Inform* 73:1–23, 2004