Vis Comput
DOI 10.1007/s00371-016-1232-1

ORIGINAL ARTICLE

Retrieval of clothing images based on relevance feedback with focus on collar designs

Honglin Li[1] · **Masahiro Toyoura**[1] · **Kazumi Shimizu**[1] · **Wei Yang**[1] · **Xiaoyang Mao**[1]

Abstract The content-based image retrieval methods are developed to help people find what they desire based on preferred images instead of linguistic information. This paper focuses on capturing the image features representing details of the collar designs, which is important for people to choose clothing. The quality of the feature extraction methods is important for the queries. This paper presents several new methods for the collar-design feature extraction. A prototype of clothing image retrieval system based on relevance feedback approach and optimum-path forest algorithm is also developed to improve the query results and allows users to find clothing image of more preferred design. A series of experiments are conducted to test the qualities of the feature extraction methods and validate the effectiveness and efficiency of the RF-OPF prototype from multiple aspects. The evaluation scores of initial query results are used to test the qualities of the feature extraction methods. The average scores of all RF steps, the average numbers of RF iterations taken before achieving desired results and the score transition of RF iterations are used to validate the effectiveness and efficiency of the proposed RF-OPF prototype.

Keywords Content-based clothing image retrieval · Collar design · Feature extraction · Saliency map · SIFT · Relevance feedback · Optimum-path forest

1 Introduction

As e-commerce continues to gain momentum in the market, more and more consumers are searching online shops for clothing items. However, sifting through the massive amounts of available products to find an item that suits one's tastes and preferences can be an arduous, time-consuming task. Many sites support keyword-based searches, but items in online shops often lack specific design-related tags and include technical names that few shoppers are familiar with. One image search approach that researchers have proposed as visual query-based alternatives to keyword-driven searches is the content-based image retrieval (CBIR) method [1], which involves expressing image content in feature vectors and then comparing the similarities between various images. Although CBIR methods eliminate the need to build queries out of keywords or other linguistic information and allows users to search for visual information with visual information input, their efficiencies largely depend on the quality of the feature vectors, and it remains to be challenged to extract the image features capturing the design of clothing well. Some researchers [2] have attempted to use color and texture, but few researchers have delved into the possibilities of developing feature vectors to capture the designs of clothing in detail.

The main contribution of this paper is twofold. One is a novel method for extracting the detailed design features of collars from 2D clothing images. The other is a prototype of clothing image retrieval system based on Relevance Feedback approach and Optimum Forest algorithm for allowing users to find clothing image of preferred design. We focus on collars because it is a well-known fact among garment designers that collar is a crucial part of a garment as it serves as the frame for one's face [3–5]. Since face is the most important visual attribute for characterizing a person, the design of a collar can largely affect the look of a garment and the overall impression of a person. In addition, the artistic formation of a collar is usually the most eye-catching part when people look for clothing due to its horizontal perspective position clos-

✉ Xiaoyang Mao
mao@yamanashi.ac.jp

1 University of Yamanashi, Yamanashi, Japan

Published online: 05 May 2016

est to the observers' eyes [6]. Although computer-assisted 3D garment design technologies [7,8] and virtual fitting systems [9,10] have become available recently, E-commerce websites for clothing shopping mainly show 2D clothing images for users to choose. Our proposed method allows users to retrieve the clothing of preferred design from such online shopping sites first, which will narrow down the candidates and speed up the procedure of virtual fitting [10].

In our study, we consider collar spread, turnover and the front design as three important design factors of a collar and develop new methods for automatically extracting those features from clothing images. For the turnover and front design, three different feature vectors based on SIFT feature and Saliency Map are designed. We further apply optimum-path forest algorithm to perform image search procedures. By incorporating the relevance feedback approach into the OPF process [11], our method enables users to search for images based on his or her subjective preferences on collar design. To validate the above proposed ideas and methods, a series of experiments are conducted. The experiment results presented in [12] could not provide sufficient validation to the proposed methods. First, only four subjects were involved in the experiments and thus the results might not be reliable enough. Second, neither the effects of the different collar types on the evaluation scores nor the bias caused by the difference in the number of images of different collar types was considered. Third, the improvements of the RF-OPF prototype on the three feature extraction methods were not discussed. In this work, we redesign the experiments and succeed in achieving reliable and detailed results to validate the effectiveness and efficiency of the proposed methods for all referred collar types.

2 Related research

The ongoing spread of e-commerce, among other factors, has prompted numerous researchers to explore the possibilities of applying search methods to clothing. Liu et al. [13], for example, proposed a method that makes it possible to use snapshots to search for clothing available in online marketplaces. The method developed by Liu et al. involves taking a picture of a person's body, separating the body into its constituent parts (feet and legs, for example) and determining the features of each part to enable users to search for images on a fashion website. Bossard et al. [14] proposed a method for classifying apparel in photographs. Using SVM and random forest allows their method to establish clothing categories like long skirts and coats and classify clothing according to sleeve length, material and other attributes, but these classification schemes were the ultimate purpose of the research; Bossard et al. did not include the idea of searching for clothing based on specific design elements. A project by Hsu et al. [2] used images with uniform backgrounds as queries for retrieving a limited scope of clothing items that one might find in an online shop. With a piece of clothing serving as the input for the method, their approach involved comparing items based on the features—color, texture, SIFT features and outline—that the pixels in the clothing regions of the given images form. features were also used in our study, which aimed to extract the designs and other characteristics of collars. None of these existing methods are capable of searching for detailed information that collar designs represent. One recent study proposed the idea of using sketches to search for clothing items with the desired design [15]. However, the method presents problems for people who lack sufficient sketching skills.

In the CBIR field, meanwhile, relevance feedback (RF) method has been drawing substantial attention for its use of dialogic feedback between the users and the system for learner-driven learning and searching purposes. Searching based on relevance feedback method makes it possible to update classifiers by showing results to users. Researchers have already tested this approach in searching for images with ambiguous thematic content, such as ocean scenes, cats and sunsets. One study has proposed a method that produces high-quality results via minimal amounts of feedback by incorporating different types of classifiers and reusing past classification results [16]. By employing a prototype with the RF approach for learning and the OPF algorithm for searching procedures, our method enables users to search for collars that align with their personal preferences, which may have positive effects on choosing the whole clothing.

3 Collar design

Generally, image searches operate on the similarity of visual attributes of multiple images. Finding a collar that suits one's tastes, however, would require the matching of design elements in greater detail. For our study, we begin by interviewing instructors at fashion colleges about the design elements of collars. Then we take their suggestions into consideration to design feature vectors that can enable fine-tuned search functionality. Clothing experts have suggested that collars generally come in the following ten types (Fig. 1) [4].

Three important elements should be taken into consideration in describing the collar designs. The first is the *collar spreads*—the ways they open. The second is *Turnover*. Figure 2 shows two examples of the collar design with turnover. The existence and shape of the turnover are some of the most distinctive features affecting the preferences of many consumers looking for buying clothing. The third is *front designs*, which refers to the types of ribbons and frills as shown in Fig. 3.

Fig. 1 Collar types

Fig. 2 Turnover collars

Fig. 3 Collar with front designs

Collar design preferences vary considerably according to buyers' tastes and needs. A person looking for work clothing, for example, would probably prefer a simple, clean look to a busy, loud design. On the other hand, a buyer trying to find something flashier to wear for a fancy occasion might opt for an ensemble that features frills or a gather. Personal tastes affect people's clothing choices in a wide variety of other ways, for instance, the likelihood of a person with a reserved, quiet personality buying a frilly design is rather slim. In hopes of enabling a search approach that reflects user preferences, we design feature vectors to describe detailed collar designs and build a search system using the relevance feedback method based on OPF algorithm. For the preliminary feature extraction process, our method involves obtaining feature vectors describing three design elements: collar spread, turnover and front design-for each clothing item in the clothing database. Three kinds of feature extraction methods based on SIFT Saliency and Saliency & SIFT, respectively, are implemented. In the runtime phase, OPF algorithm is used to classify the images into relevant or irrelevant based on the initial training images. The OPF classifier is refined through iterative relevance feedback from the users. The framework of the proposed approaches is shown in Fig. 4.

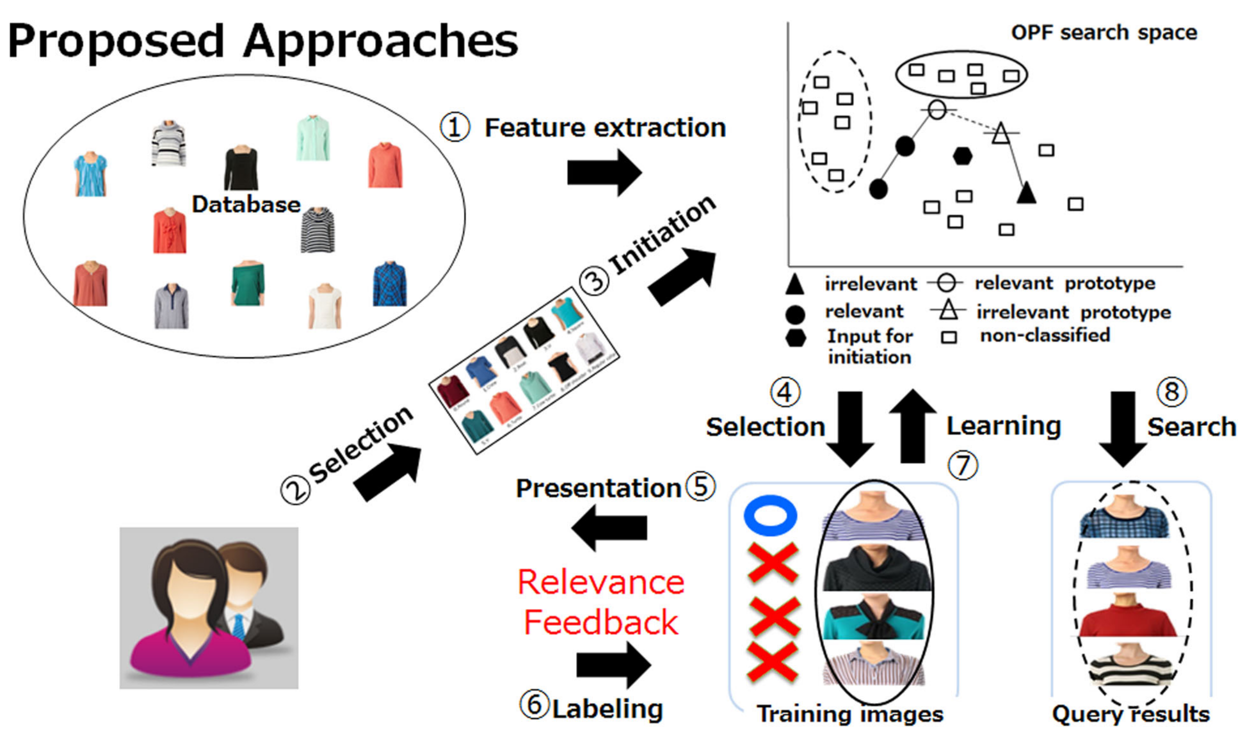

Fig. 4 The framework of the proposed approaches

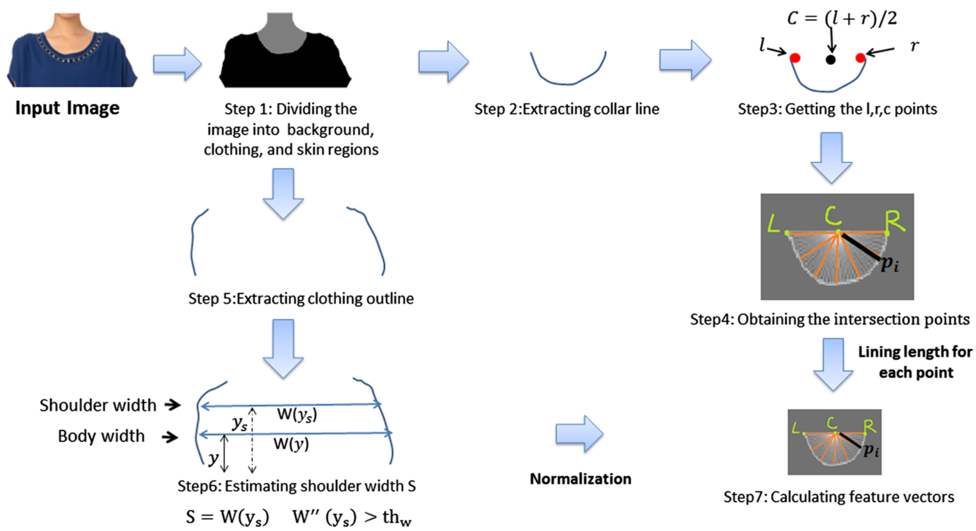

Fig. 5 Computing the feature vectors of the collar spread

4 Designing and extracting feature vector

In order to adopt the RF approach and use the OPF algorithm to search the items of preferred collar design, it is necessary to build a feature vector space that effectively reflects the design features of collars. To describe the three important elements, which are spread, turnover and front designs, the feature vectors are designed and extracted as below.

4.1 Collar spread

For our study, we limit our scope to images of tops that showed the wearer's upper body only. We also assume that each image has a monochromatic background showing a human subject from the front and the color of which is at least somewhat different from that of the subject's skin. Incorporating more advanced image processing technologies would make it possible to ease these restrictions, but doing so would deviate from the main focus of our study. Figure 5 illustrates the process of extracting feature vectors that represent the collar spread (opening). We use the following steps to extract features:

Step 1: Dividing the image into background, clothing and skin regions.

This step is the preparation work for identifying the collar line (the clothing-skin boundary) of Step 2 and determining the outline of the clothing in the image (the clothing-background boundary) of Step 5. The collar spread features can be expressed according to the depth value for each angle emanating from the subject's neck, as indicated by the result of Step 7 in Fig. 5.

We separate the background and the foreground of the input image. To extract the skin color region, we use skin color S, the dominant color of the area between the neckline and the chin, as our learning data and extract the pixels with the shortest Mahalanobis' generalized distances from skin color S and use their color t as the skin color. And thus we can define the collar line and clothing outline by locating the skin, clothing and background boundaries. We place single seeds in the skin, clothing and background areas respectively and then watershed algorithm was used to divide the image into the three regions based on these seeds.

Step 2: Extracting collar line

With the image divided into three regions, we then extract pixels from where the skin and clothing regions meet to identify the collar line.

Step 3: Getting the l, r, c points

This step involved extracting the coordinates of the left end 1, center c and right end r of the collar line, which will be used in Step 7 to compute the feature vectors representing the spread of collar. We use chain codes to trace the contour of the entire collar line identified in Step 2, which allow us to determine the coordinates of left end 1 and right end r. Then the coordinates of 1 and r are used to determine the center c of the collar.

Step 4: Obtaining the intersection points

Then we extend radial lines from the center c in the direction of $\alpha_i = i \times \pi/36$ $(i = 0, \ldots, 36)$ until the lines intersect with the collar line that we extracted in Step 2. The intersections are defined as points p_i $(i = 0, \ldots, 36)$.

Step 5: Extracting clothing outline

With the image divided into three regions in Step 1, we extract pixels from where the clothing and background regions meet to generate the clothing outline.

Step 6: Estimating shoulder width S

Using the clothing outline founded in Step 5, we next calculate shoulder width. Its value serves as the basis for normalizing the feature vector calculations in Step 7. By

Fig. 6 Comparison of the result by spread feature vectors on clothes with *V* opening, square opening, turn over and front design collar

normalizing the feature vectors, we ensure that the feature quantities obtained are independent of image size.

To calculate shoulder width, we use the differential of body width. Denoting the body width at height y in the clothing outline as $W(y)$, the shoulder width S is given as $W(y_s)$ at the height y_s, where the second-order derivative of $W(y)$ exceeds a given threshold.

$$S = W(y_s) W''(y_s) > \mathrm{th_w} \tag{1}$$

Step 7: Calculating feature vectors

With the intersection points p_i $(i = 0, \ldots, 36)$ obtained in Step 4 and the center c obtained in Step 3, the lengths $|c - p_i|$ $(i = 0, \ldots, 36)$ are computed and normalized based on shoulder width obtained in Step 6. Finally, the normalized values constitute the 36d feature vector of the collar.

Figure 6 shows the searching results using the above spread feature vector only. We can see the collar spread feature is well captured (the first two rows), but the feature of turnover and front design cannot be distinguished (the last two rows).

4.2 Turnover and front design

To capture the features of turnover and front design, we design three different feature vectors, which are described in the following as SIFT-Based Feature (SBF), Saliency-Based Feature (SABF) and Saliency & SIFT-Based Feature (SSBF).

SBF

SIFT is the most commonly used size and orientation invariant feature. Using a gradient histogram around arranged points makes it is possible to capture local details of image contents. It can be expected to capture ribbon shapes and other design details. To accelerate the feature matching in image retrieval, we combined the bag-of-visual-word method [17–19] with the SIFT feature extraction. First we compute the 128d local SIFT feature. Then K-means method is used to cluster the SIFT feature vectors into 500 clusters (500 here is empirically given). The centers of clusters are used as codewords. Finally, each SIFT feature vector is mapped to the closest codeword to obtain the histogram of the codewords, which is a 500d feature vector. We compute the above feature vector for the region from the collar to the chest (Collar-SIFT) and the whole image (Whole-SIFT).

SABF

Figure 7 illustrates the comparison of searching results between a piece of plain front design clothing and a striped V collar one using SIFT feature. While this approach manages to find relatively good matches (the first row) for the plain clothing image, it fails to produce the same quality results (the second row) for a striped one. The second row results shows that images with similar clothing textures—not collar designs—appear in the top search results. In other words, the SIFT features can be too dominated by texture features.

Fig. 7 Comparison of search results between a piece of plain clothing and a striped one with Sift feature

In order to reduce the effects of the texture features on the collar designs, we propose a new feature extraction method using saliency map proposed by Xiaodi et al. [20]. Drawing on the mechanisms of human visual attention, the saliency map method allows users to determine the area of an image that observers are most likely to focus on. A saliency map calculates the degree of attention not based on the presence of patterns, but on the differences between a given location of an image and its surroundings. This method can be expected to be effective for obtaining features from the front designs as well as turnover, which are the decorative embellishments serving for attracting attention. As long as the design is distinct in some way from its surroundings, the saliency map method can get good matches even if the clothing consists of textures.

Figure 8 shows the results from two examples consisting of textures. The frill area of the example shown in Fig. 8a exhibits prominent differences from its surroundings on the saliency map. Even though the clothing consists of large checked textures, the turnover can still be captured by the Saliency Map (Fig. 8b).

The appearance of a front design, meanwhile, depends heavily on the size, length and breadth of their attention-drawing elements. Various shapes and sizes of ribbons and frills have an impact on personal preference. Frills that cover a considerable area on a piece of clothing, for example, create quite a different visual impression from minimal, dainty frills on the top of the collar. One can also locate areas with the highest concentrations of elements that diverge from the general look of a given clothing item. The frills in the images of Fig. 9a, for example, spread out across the width of the wearer's chest to create a relatively showy impression. The ribbons in the images of Fig. 9b, meanwhile, give the clothing a more extravagant but vertically oriented appearance. To capture these design-related differences, the proposed method lays a given grid (3×5) over the saliency map as shown in Fig. 9b. We obtain a 15d feature vector, where the pixels values' sum in each grid cell represents a dimension of the feature vector.

SSBF

This approach, however, can capture the overall shape of front design only. If the clothing in the input image features a small group of frills, the design details of the frills cannot be captured. We thus propose a method that combines the saliency map and SIFT feature. Fortunately, Saliency Map proposed by Xiaodi et al. [20] allows controlling the level

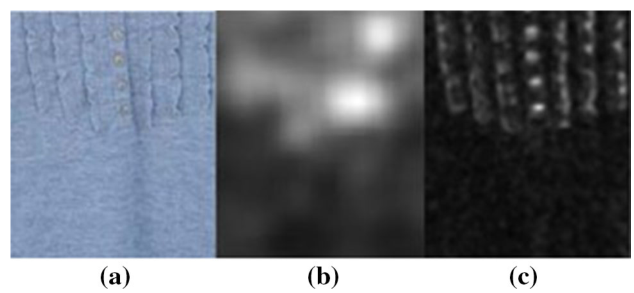

Fig. 10 Saliency map computed with the method given in [20]. **a** Input image. **b** Saliency Map computed using low frequency band. **c** Saliency Map computed using high-frequency band

Fig. 8 The results of Saliency Map acting on two clothing images. **a** An example with front design. **b** An example with turnover

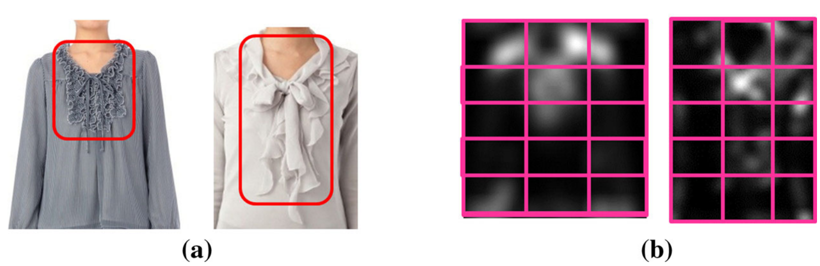

Fig. 9 Feature vector design for capturing front designs. **a** Examples of ribbons. **b** 3×5 grid arranged on saliency map for capturing the spatial distribution of front designs

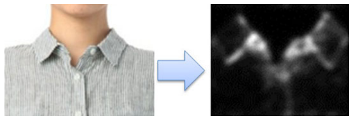

Fig. 11 Extracting turnover features in a high-detailed Saliency Map

Fig. 12 Results by SBF (*top*) and SSBF (*bottom*)

of detail. Using the high-frequency band to compute the difference from the average, we obtain results that retain the frill details (Fig. 10c). Figure 11 shows the result of using a Saliency map over a striped shirt with a turnover collar. If we are to apply normal edge detection to this image, with its prominent pattern, it will be difficult to determine where the turnover is. Using a Saliency Map with the proper level of detail, however, allows us to limit the impact of the pattern to a certain degree. Then by computing the SIFT features from the saliency map image, we are able to obtain the details of collar design while eliminating the influence of texture features of the clothing item.

In Fig. 12, the images of the top row are the search results obtained using SIFT feature only. The clothing items with round neck are also included in the results as those items have the gather pattern similar to the strip texture of the input clothing. Most of those items are eliminated in the results shown in the bottom row by combining Saliency Map and SIFT feature. In addition, more items with turnover collar or similar collar spread are included in the search results. However, the query results are still not good enough. So a prototype combining Relevance Feedback approach with OPF classifiers is proposed to improve the qualities of the query results as follows.

5 The RF-OPF prototype

Relevance feedback (RF) method is a critical component in our CBIR prototype, which makes it possible for users to interact with the system and thus reflects their design preference in the query. The classifiers, another critical component of our CBIR system, are used for processing queries. Their efficiency (related with response time) and effectiveness (related with users' satisfaction) are very important for evaluating the quality of this CBIR system. In our prototype, we use the optimum-path Forest (OPF) [11,21,22] classifier for query and classification. OPF works by modeling the classification as a graph partition in a given feature space. It starts as

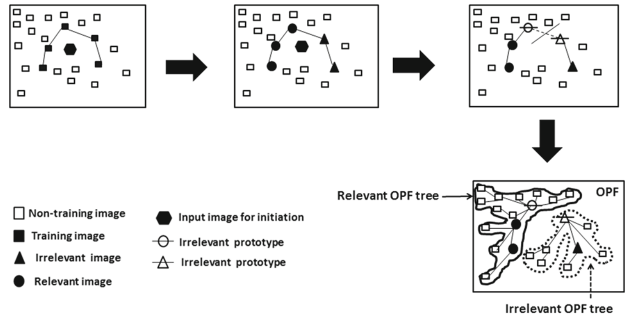

Fig. 13 Generating the OPF classifier

a complete graph, whose nodes represent the feature vectors of all images in the database. All pairs of nodes are linked by arcs which are weighted by the distances between the feature vectors of the corresponding nodes (referred as costs here and after). As illustrated by Fig. 13, given a set of training nodes, a minimum spanning tree (MST) can be generated from the complete graph. Then the adjacent training nodes are marked as prototypes if they belong to different classes, which are relevant and irrelevant in our case. The partition of the graph is carried out by the competition process among prototypes, which offer optimum paths to the remaining nodes of the graph. The optimum paths from the prototypes to the other samples are computed by the algorithm of the image foresting transform (IFT), which is essentially Dijkstra's algorithm modified for multiple sources and more general path-value functions. At last, all the non-prototypes are connected with a prototype directly or indirectly with the minimum costs. With the prototypes as the roots and the non-prototypes as the intermediate and terminal nodes, the optimum trees are built, which constitute the optimum-path forest (OPF). Compared with SVM, ANN-MLP and K-NN, OPF is usually superior to ANN-MLP and K-NN in accuracy and significantly outperforms SVM in computation time [11,21,22], which is very important in a prototype based on RF approach that generates results in a dialogic fashion.

All the images in the database are represented by the feature vectors extracted with one of the three methods referred in the Sect. 4. The first is Spread + Collar-SIFT + Whole-SIFT represented by $1036d$ ($36d + 500d + 500d$). The second is Spread + Saliency represented by $51d$ ($36d + 15d$). The last is Spread + Saliency-SIFT represented by $1036d$ ($36d + 500d + 500d$).

Based on the OPF classifier described above, we build our RF-OPF prototype, using the following steps:

1. The initial training set containing images of ten different collar types plus front design type is presented to user. When the users choose one desired collar type from the initial training set, the five images with smallest L2 distance to the chosen image in the feature spaces are returned to the users.

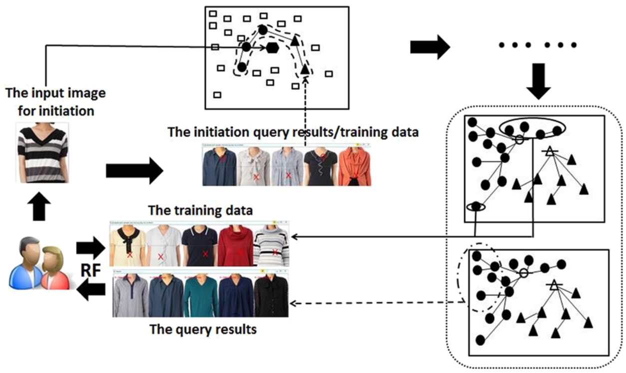

Fig. 14 The operating mechanism of the RF-OPF prototype

2. The user evaluates the results. If he/she is satisfied with the initial query result, the RF ends up without using the OPF classifier. If not satisfied, he/she should mark the images with × representing irrelevant or O representing relevant.
3. The first five images marked with × or O constitute the original training set for building an OPF classifier mentioned above. The procedure is illustrated as Fig. 14 below and the detail of the OPF algorithm was described in [21]. Then we use the OPF classifier to sort the unclassified images of the database into two classes, relevant and irrelevant.
4. The RF-OPF prototype then chooses five images with the maximum degree of relevance as the query results and five images with the minimum degree of relevance as the training samples from the relevant class and returns them to the users. The new marked training samples will be merged into the former training samples to build a new OPF classifier for the next RF phase if the users are not satisfied. This procedure continues until the users are satisfied.

The framework and operating mechanism of our RF-OPF prototype are shown in Figs. 4 and 14, respectively.

When selecting the query results and the training samples, we compare the costs of paths from all non-training images to all relevant and irrelevant prototypes. The five images which belong to relevant class with the largest ratio of costs to the relevant prototypes over costs to the irrelevant prototypes are chosen as the query results. The training samples are the five images which belong to relevant class and have the smallest ratio of costs to the relevant prototypes over costs to the irrelevant prototypes. In our implementation, the cost of the arc connecting two adjacent nodes of the OPF feature space is calculated with the L2-norm. And the cost of a path is the maximum value of the costs of all arcs constituting the path.

Assuming the number of relevant prototypes and irrelevant prototypes to be k and m and denoting the k relevant prototypes and m irrelevant prototypes as $p_i(i = 1, 2...k)$ and $q_j(j = 1, 2...m)$, we consider $k \times m$ pairs of (p_i, q_j)

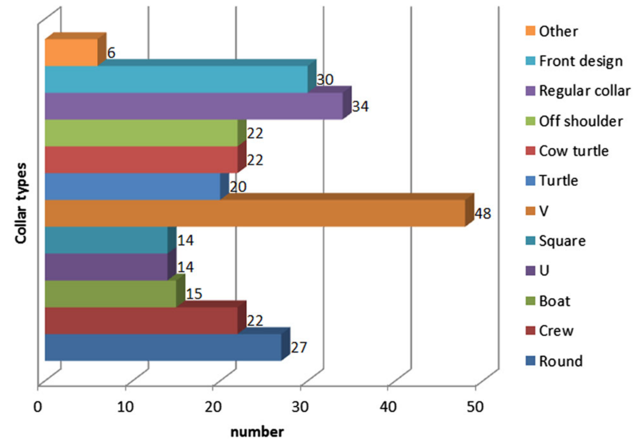

Fig. 15 Number of images for the ten collar types and clothing with front design

in computing the ratio of the costs of paths to the relevant and irrelevant prototypes. Let $\mathrm{CR}_{\mathrm{U}\to p_i}$ and $\mathrm{CI}_{\mathrm{U}\to q_j}$ represent the costs of the path from a non-training sample U to the relevant prototype p_i and the irrelevant prototype q_j, respectively. $\mathrm{Relevance}_{\mathrm{U}\to(p_i,q_j)}$, which represents the ratio of $\mathrm{CR}_{\mathrm{U}\to p_i}$ over $\mathrm{CI}_{\mathrm{U}\to q_j}$, is computed as:

$$\mathrm{Relevance}_{\mathrm{U}\to(p_i,q_j)} = \left\| \mathrm{CR}_{\mathrm{U}\to p_i} - \mathrm{CI}_{\mathrm{U}\to q_j} \right\| \quad (2)$$

We use subtraction instead of ratio to avoid encountering the overflow problem when $\mathrm{CI}_{\mathrm{U}\to q_j}$ is very small.

6 Experiment and discussion

6.1 Experiment

The images used for experiments are gathered from the Internet. All images have monochromatic background and show upper bodies from the front. Totally there are 274 images including ten different collar types and the clothing with front design as shown in Figs. 1 and 3. The number of images for each of the ten collar types and the clothing with front design is shown in Fig. 15. Because it should be easier to find a certain collar type if the number of it in the database is high, we take into consideration the proportion of image numbers of different collar types in evaluating the efficiency of proposed methods.

For all images, we pre-computed their Spread, SBF, SABF, SSBF vectors. The Spread feature vector has 36 dimensions representing the distance from the neck center to the collar line measured every $\pi/36$ as shown in Fig. 5. The SIFT feature vector is computed for the collar area and the whole upper clothing area, which consists of 500 dimensions separately (Fig. 7). When being used alone, the SABF vector

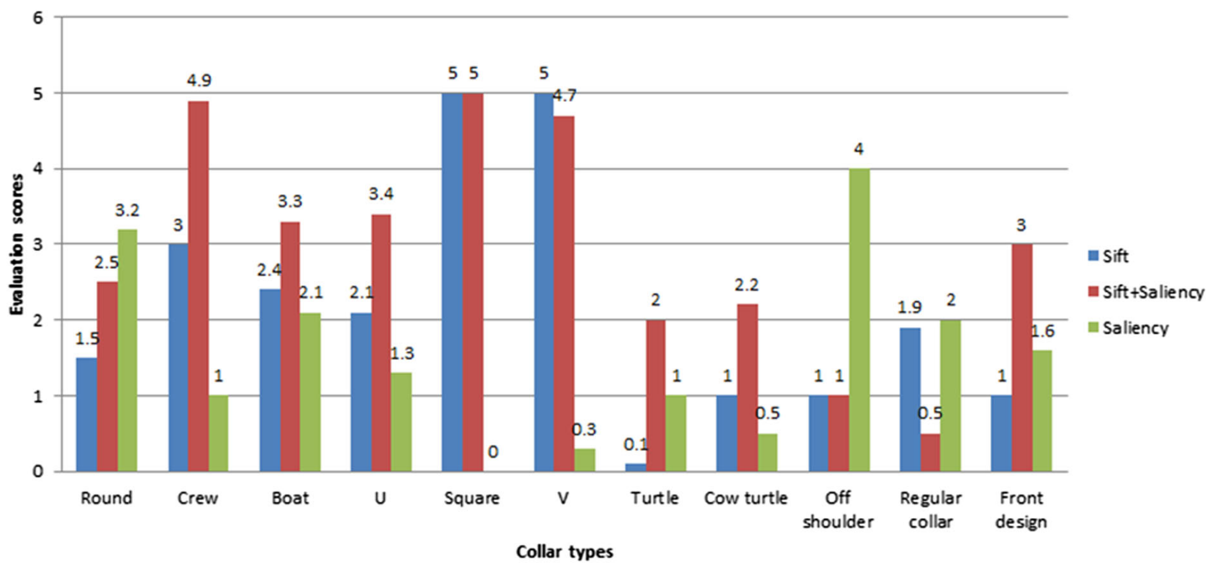

Fig. 16 Comparison of the average scores of the three methods on the collar types

has 15 dimensions (Fig. 9). When combined with SIFT, it has 500 dimensions.

As described in Sect. 5, our RF procedure starts by letting the users choose one desired image out of an initial training set containing images of ten different collar types plus front design type. This step is used to build the initial OPF classifier and has big impact on the results of the following RF steps. In other words, the type of collar the users choose at the initiation step should be an important parameter in designing the experiment for two reasons. One is its relationship with the feature vectors; the effectiveness of the proposed feature vectors should be different for different collar types. The other is its number used in the experiment as it should be easier to find a certain collar type if there are more images of such type in the database. The experiments presented in [12], however, ignored these factors completely. To solve the problems, in our new experiments, each subject is asked to test all the ten collar types plus one front designs for the three different combinations of feature vectors: Spread + SBF, Spread + SABF and Spread + SSBF. Therefore, each subject performed 33 tests in total and we compared the effects of the three feature vectors for each collar type, respectively. To eliminate the bias caused by the population of collar types in the database, the number of images of each collar type is used to weight the score inversely when compare the average score of the three feature vectors for all images. While previous experiments used only four subjects [12], we improved the reliability of experiment results by expanding the number of subjects to 10. The ten subjects are female college students from school of nursing. For each of the three feature vectors, they were asked to initiate the RF with each collar type in the initial training set. At each step of RF, the subjects were asked to mark each of the five training images as "relevant" (O) and "irrelevant" (×) and evaluate each of the query results as "satisfactory" (O) and "unsatisfactory" (×). Then a score ranging 0–5, which corresponds to the number of satisfactory images, is automatically computed for the query results of each step.

6.2 Comparisons of feature vectors (Spread + SBF, Spread + SABF, Spread + SSBF)

Figure 16 shows the evaluation scores (averaged by ten subjects'scores) of the initial query results for the three feature vectors. We can observe that SBF and SSBF perform well at the first six collar types, especially for the Square and the V types. SSBF works well for the Crew type also and SABF is excellent for the Off shoulder type. We use the weighted (inversely proportional to their number percentage of total) average scores of the collar types to compare the overall qualities of the three methods, which will reduce the bias caused by the different number of images of different collar types. As shown in Fig. 17, the weighted score of the SSBF method is higher than those of the other two. One-tailed paired t-test reveals that the average score of SSBF in Fig. 17 is significantly higher than that of SBF and SABF at significance level 5 % ($p = 0.05$).

6.3 Improvement by RF-OPF

Although the results shown in Fig. 17 demonstrate that SSBF is more effective than the other two feature vectors, the initial query results of each feature vector is not good enough. It is because that the weighted average score of even the best case(SSBF) is just 3.08 against the full score of 5. We expect

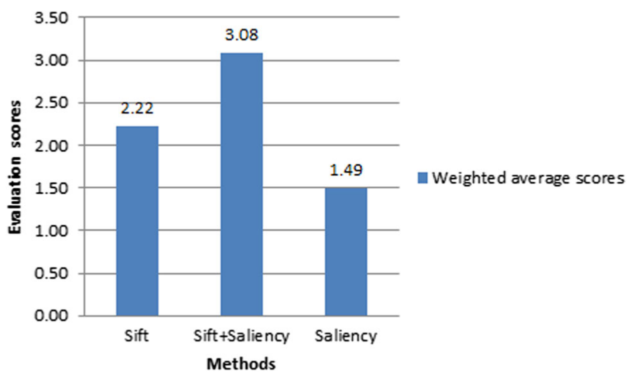

Fig. 17 Comparison of the weighted average scores of the three methods on the collar types

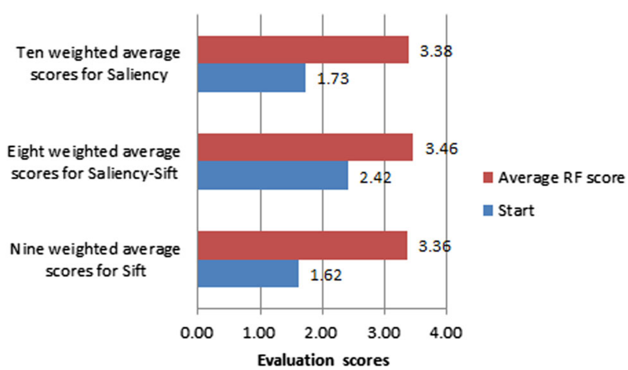

Fig. 18 The improvement of the RF-OPF prototype on the methods

that continuing the RF iterations illustrated in Sect. 5 can further improve the qualities of the query results.

So if the subjects get few satisfactory images from the first query results, they are asked to continue the query and evaluation process until they are satisfied with the number of the satisfactory images returned by the RF-OPF prototype. The evaluation scores at each RF step are also averaged by the ten subjects and inversely weighted by the number of collar types as we have done in dealing with the initial query results. Since the subjects always terminated the RF when the query results contain four or five desired images (the evaluation scores are correspondingly 4 or 5), the times of RF iteration they took also reflects the efficiency of our RF-OPF prototype. To validate whether RF-OPF is useful for improving the query results, we compare the average evaluation scores of all RF steps with the first query scores. Please note that not all collar types are used in computing the average evaluation scores of all RF steps. It is because the results of the Square and the V collar types with the SBF and the SSBF methods satisfy the subjects at the initial query phase, and the Crew collar type with the SSBF method satisfies the subjects at the initial query phase too. On the other hand, Square collar type with the SABF cannot get any suitable image at the initial query phase, which generates a score of zero that cannot be compared in this experiment. Therefore, as is shown in the left title of Fig. 18, there are ten, eight and nine collar types with the above three feature vector combinations to be compared. Figure 18 shows that the RF-OPF prototype makes great improvement on the initial query results for all the three feature vectors, which are 107.69, 43.28 and 95.47 %, respectively. But for those collar types not involved in the RF phases, the effectiveness of the RF-OPF prototype cannot be validated in this way. We partially address this problem with another experiment described later.

The average RF iteration times the subjects take to reach their objects are the evidences to validate the efficiency of the RF-OPF prototype directly and the corresponding feature vectors indirectly. We compare the average RF times of the three methods from the ten subjects on each collar type in Fig. 19. From this figure, some useful information about the efficiency of the RF-OPF prototype on the collar types can be observed, such as the Square collar type being still very

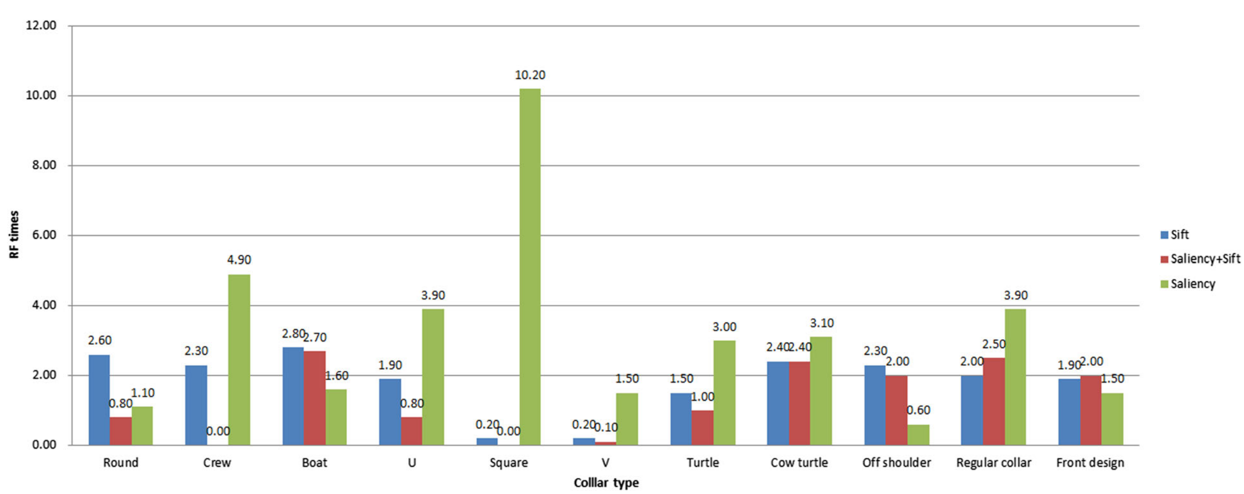

Fig. 19 Comparison of the RF times of the three methods on the collar types

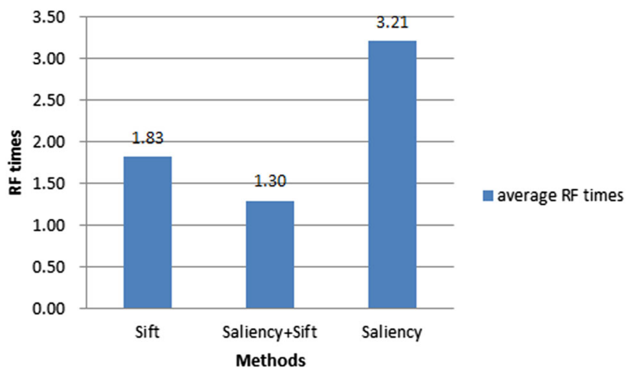

Fig. 20 Comparison of the average RF times of the three methods

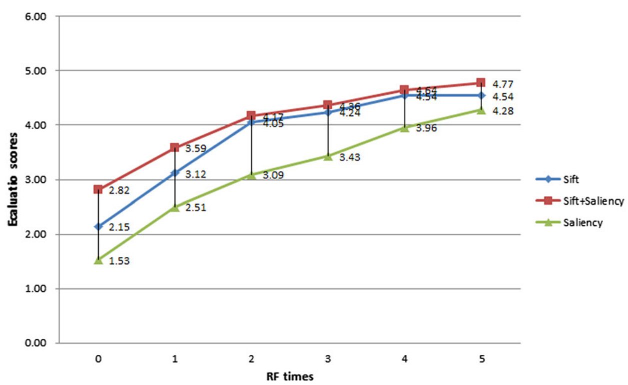

Fig. 21 Scores trends of five RF steps on the three feature vectors

difficult to be found with the SABF vector even with the help of RF-OPF prototype. Figure 20 shows that the SSBF vector needs the least average RF times (1.30) to make the users satisfied, which demonstrates its efficiency. A one-tailed paired t test for the RF times in Fig. 20 shows that the difference between the SBF and SSBF and that between the SBF and SSBF are statistical significance ($p = 0.05$). The difference between the SBF and SABF is statistical significance at some extent ($p = 0.1$).

Since the experiment results of Fig. 18 do not contain the entire collar types, the effectiveness of the RF-OPF prototype cannot be validated completely. To solve this problem, we ask all the subjects to perform five times RF steps no matter what the results are at each step. Since the highest average RF times of the three feature vectors shown by Fig. 20 is 3.21, 5 times is considered to be enough for this experiment. With the evaluation scores of the initial query and the five RF steps' results, we get Fig. 21. It shows that the RF-OPF prototype improves the first query scores of the three feature vectors from 2.15, 2.82 and 1.53 to 4.54, 4.77 and 4.28, respectively. The improvements are significant and the scores trends are incremental in general.

At last we use the SSBF feature and the RF-OPF prototype to query similar collar images for the two given images

Fig. 22 **a** The top five results for a striped V collar image with the SSBF method by the second RF phase. **b** The top five results for a front design image with the SSBF method by the second RF phase

mentioned in Fig. 7. The results are shown in Fig. 22, which shows that the striped V collar type image can get five images of similar collar design at the second RF step and the image with front design can get four images of similar design at the second RF step too. Note that, although the initial query results or even the first RF results may be very unsatisfactory, we can always get better results after several RF iterations.

7 Conclusion

Collar design plays an important role when people choose the desired clothing. In this paper, we focus our research on the retrieval of clothing image based on the collar design. Three different feature vectors SBF, SABF and SSBF combined with Spread for capturing the detailed features of collar design and a CBIR prototype based on RF method and OPF classifiers are proposed. Through experiments, it is proved that SSBF is the best among the three feature vectors in terms of both effectiveness and efficiency. Our experiment results also demonstrated that the proposed RF-OPF prototype improves the qualities of the query results significantly.

But there are still some technical problems to be solved. 1. The SSBF feature vector still does not react well to some collar types, such as the Off shoulder and the Regular collar types. 2. The subjects may be confused by some similar collar types, such as the Round and the U collar type, the Crew and the Boat collar type during the RF phases in the experiment. In addition, the collars combing multiple design features, such as a collar combing turnover with V spread or front design, are difficult to be classified by the subjects. 3. The score of the query results is computed as the number of satisfactory images without considering the degree of satisfaction of each image in the query results. To solve the first problem, we plan to develop and experiment with other new feature vectors for capturing the collar design. The second problem can be solved by improving the spread feature vectors for better describing the roundness, width and angle, so as to better discriminate collar types of similar shape, like the Round, Crew, Boat and U collar types. The 3rd problem can be solved by asking subjects to score each of the retrieved images and analyze the trend of highest and average scores of RF iterations. The current image database is relatively small. We need to gather more images for improving the reliability of the experiment results. Another important issue is that the initial image sets for letting user to select one collar type to initiate the RF have large impact on the results of succeeding steps. We need to explore some new methods which can always avoid misleading the building of the classifier.

3D garment design is a rapidly developing field [7,8]. Presenting 3D clothing images in clothing-shopping websites will help people find the desired clothing more easily. So it is very important for us to extend our research to deal with 3D collar design features in the future. Although collar is very important for choosing clothing, Color, texture and the design of other parts of garments also affect how people choose clothing at different extent. We are going to extend our retrieval system by considering other features. It should be easy to capture color and texture features with existing computer vision technologies. The technology provided in [13] can be employed for capturing the overall shape of garments. Moreover, the proposed RF-OPF system should have high potential to deal with more design factors as it is reported that OPF is superior to conventional learning algorithms such as ANN-MLP, K-NN and SVM in both computation time and accuracy, especially in complex situations, i.e., with a large amount of overlapped regions [21,23].

Acknowledgments We would like to thank Prof Takami Yamamoto for her valuable comments. This work was partially supported by JSPS KAKENHI (16H05867, 25280037).

References

1. Veltkamp, R.C., Tanase, M.: Content-based image retrieval systems: a survey. Technical Report UUCS-2000-34, Utrecht University, The Netherlands (2000)
2. Hsu, E., Paz, C., Shen, S.: Clothing image retrieval for smarter Shopping. EE368, Department of Electrical and Engineering, Stanford University (2011)
3. Hackler, N.: UK Cooperative Extension Service. University of Kentucky-College of Agriculture, CT-LMH, pp. 1–9 (1997)
4. Onuma, J.: Coordinate Technique-Appealed, 1st edn. Bunka Publishing, Bunka Fashion College (2001)
5. Pundir, N.: Fashion Techology-Today and Tomorrow. A Mittal Publication, New Delhi (2007)
6. Wang, L., Tian, B.: An analysis of factors determining the shape of collar. J. Panzhihua Univ. **2**, 87–89 (2008)
7. Fang, J.: 3D collar design creation. Int. J. Cloth. Sci. Technol. **15**(2), 88–106 (2003)
8. Liu, Y., Zhang, D., Yuen, M.: A survey on CAD methods in 3D garment design. Comput. Ind. **61**(6), 576–593 (2010)
9. Zhang, X., Wong, L.Y.: Virtual fitting: real-time garment simulation for online shopping. In: ACM SIGGRAPH Posters (2014)
10. Hauswiesner, S., Straka, M., Reitmayr, G.: Virtual try-on through image-based rendering. IEEE Trans. Vis. Comput. Graph. (TVCG) **19**(9), 1552–1565 (2013)
11. Silva, A.T., Falcao, A.X., Magalhaes, L.P.: Active learning paradigms for CBIR systems based on optimum-path forest classification. Pattern Recogn. **44**, 2971–2978 (2011)
12. Shimizu, K., Yang, W., Toyoura, M., Mao, X.: Relevance feedback based retrieval of cloth image with focus on collar design. In: 2015 International Conference on Cyberworlds, pp. 137–144. IEEE, Visby (2015)
13. Liu, S., Song, Z., Liu, G., Xu, C., Lu, H., Yan, S.: Street-to-shop: cross-scenario clothing retrieval via parts alignment and auxiliary set. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3330–3337 (2012)
14. Bossard, L., Dantone, M., Leistner, C., Wengert, C., Quack, T., Gool, L.V.: Apparel classification with style. ACCV12 **4**, 321–335 (2012)
15. Kondo, S., Toyoura, M., Mao, X.: Skirt image retrieval based on sketches. Sketch-based interfaces and modeling, pp. 11–16 (2014)
16. Tekawa, M., Hattori, M.: Improvement of reuse of classifiers in CBIR using SVM active learning. Proc. ICONIP **2**, 598–605 (2010)
17. Fei-Fei, L., Fergus, R., Torralba, A.: Recognizing and learning object categories. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, short course (2007)
18. Qiu, G.: Indexing chromatic and achromatic patterns for content-based color image retrieval. Pattern Recogn. **35**(8), 1675–1686 (2002)
19. Nister, D., Stewenius, H.: Scalable recognition with a vocabulary tree. In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, pp. 2161–2168 (2006)
20. Hou, X., Zhang, L.: Saliency detection: a spectral residual approach. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
21. Papa, J.P., Falcao, A.X., Suzuki, C.T.N.: Supervised pattern classification based on optimum-path forest. Int. J. Imaging Syst. Technol. **19**(2), 120–131 (2009)
22. Silva, A.T., Falcao, A.X., Magalhaes, L.P.: A new CBIR approach based on relevance feedback and optimum path forest classification. J. WSCG **18**(1–3), 73–80 (2010)

23. Papa, J.P., Falcao, A.X.: Optimum-path forest: a novel and powerful framework for supervised graph-based pattern recognition techniques. Institute of Computing University of Campinas, pp. 41–48 (2010)

Honglin Li received the B.Sc. degree in Automation from Central South University and M.Sc. in Computer Science from Huaqiao University in 2002 and 2012 respectively in China. He began to study as a Ph.D candidate in Computer Science in University of Yamanashi from 2015. His research interests include computer vision, pattern recognition and data mining.

Masahiro Toyoura received the B.Sc. degree in Engineering, M.Sc. and Ph.D. de-green in Informatics from Kyoto University in 2003, 2005and 2008 respectively. He is currently an Assistant Professor at Interdisciplinary Graduate School, University of Yamanashi, Japan. His research interests are augmented reality, computer and human vision. He is a member of ACM and IEEE Computer Society.

Kazumi Shimizu received her B.Sc. degree in engineering from computer media department, University of Yamanashi. She is currently a master student at University of Yamanashi. Her research interests include image processing and machine learning.

Wei Yang received the B.Sc. degree in Engineering, M.Sc. and Ph.D. degrees in Computer Science from University of Yamanashi in 2010, 2012and 2015 respectively. Her re-search interests include non-photo realistic rendering and computer aesthetics.

Xiaoyang Mao received her B.Sc. in Computer Science from Fudan University, China, M.Sc. and Ph.D. in Computer Science from University of Tokyo. She is currently a Professor at Interdisciplinary Graduate School, University of Yamanashi, Japan. Her research interests include texture synthesis, non-photo-realistic rendering and their application to scientific visualization. She is a member of ACM and IEEE Computer Society.