Sketch Based Skirt Image Retrieval

Shin-ichiro Kondo* Masahiro Toyoura Xiaoyang Mao
University of Yamanashi

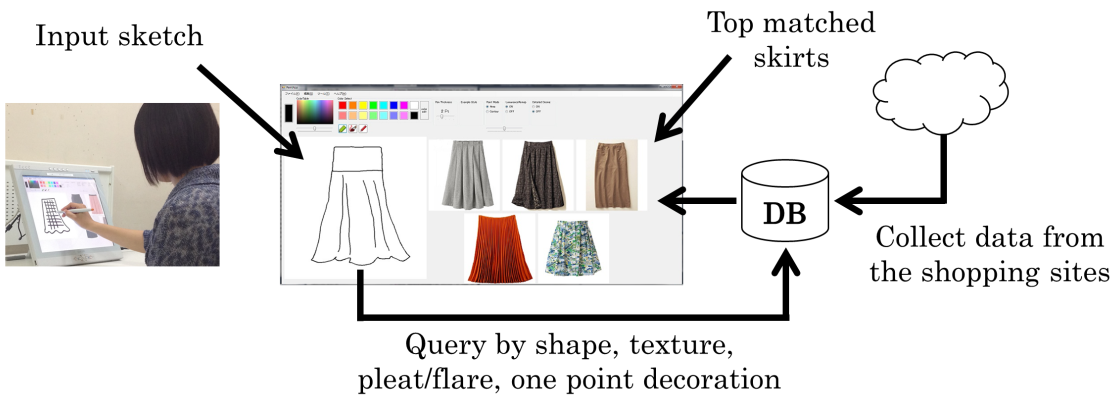


Figure 1: *Overview of interactive, skirt retrieval system based on a sketch.*

Abstract

Although many online shops allow users to search for clothes by categories or keywords, it is usually impossible to specify the details of the design. This paper presents a new technology for retrieving skirt images based on sketches. We first conducted a user study to investigate the typical features illustrated in a sketch. Then algorithms have been developed for automatically extracting those features from both the skirt images and the sketches. A prototype system has been implemented to retrieve and present the best matched skirts in real time when a user interactively sketches her imagined skirt on the canvas.

CR Categories: I.3.4 [Computer Graphics]: Picture description languages— [I.4.7]: Image Processing and Computer Vision—Feature Measurement;

Keywords: Sketch-based image retrieval, feature extraction

1 Introduction

Fashion sites have spread rapidly on the Internet. It has become possible to buy clothes easily from online shops. However, it is usually difficult to find the clothes we want from a wide selection. When we buy clothes, we often have some ideas about the design we want in advance. Although many online shops allow users to search for clothes using keywords, usually the number and categories of the keywords that can be used are limited, and it is impossible to specify the details of the design. Recently, several techniques have been developed for retrieving clothes using photographs [Liu et al. 2012; Yamaguchi et al. 2012]. However, it is not always the case that a user would have photos of the clothes that he or she wants. Additionally, the design one wants may be inspired by several photographs.

On the other hand, sketches are known to be effective for representing the ideas that humans have in mind. This study proposes a new approach that employs sketches for retrieving clothes of desired designs. To establish the framework of the search system, we focus on the search for skirts. We conducted preliminary experiments to find the important features in retrieving skirts, and we propose a method for extracting those features robustly from both the skirt images and the sketches. We have implemented a prototype system (as shown in Figure 1) for sketch-based, interactive, skirt image retrieval and verified the results through experiments.

2 Related work

Liu et al.'s recent work has presented a technology for retrieving clothing images from fashion sites, based on the photographs given by the user [Liu et al. 2012]. By using the different parts of the human body as the cue, it enables the user to robustly query the clothes in a photograph, even if the photograph has a complex background. The technology employs a detector to detect the different parts of the human body and divide the upper body into 20 sections and the lower body into 10 sections. Then features are calculated for each section and compared with the images in the fashion sites. As a similar study, Yamaguchi et al. have considered the human posture for parsing clothes in fashion photographs [Yamaguchi et al. 2012]. They first divide an image into regions corresponding to the skin, clothes, and so on, by segmenting the image into small homogeneous regions, called superpixels, using attributes such as brightness or color. Then a pose estimator is employed to estimate the body posture, which is used for labeling the clothes. They estimate the attributes of the garment by computing the co-occurrence probability of the different parts of the photo. For example, the shirt should be near the jacket but not near the shoe. Although the two cited state-of-the-art technologies make it possible to retrieve clothes based on photographs, it is not always possible for a user to have the photograph of the wanted garment ready before the re-

*e-mail:g13mk011@yamanashi.ac.jp

SBIM 2014, August 08 – 10, 2014, Vancouver, British Columbia, Canada.

ACM 978-1-4503-3018-3/14/08 $15.00

trieval. Moreover, in many cases, a user may want to search for a garment whose design combines the designs shown in several photographs.

Tseng et al. have also adopted sketches for clothing retrieval [Tseng et al. 2009]. Their work mainly focuses on the shapes of the clothes and proposes a retrieval technique that is robust to the changes in geometry. An existing segmentation algorithm called J segmentation is used to remove the influences of wrinkles, textures, and shadows. Segmentation plays an important role in achieving a good retrieval result. If the clothes have a lot of wrinkles, especially near the contours, their algorithm fails to produce reasonable results. Belongie et al. have used the shape context [4] as the shape descriptor and the Harris corner detection for selecting the sample points for the shape context. They have evaluated their technique with a small dataset consisting of around 90 items, and the average success rate is below 50%. Moreover, since the technique essentially performs shape matching only, it cannot deal with other features such as the texture, which is one of the most important factors of design.

Drawing a sketch can be a difficult task for a novice user. Lee et al. have proposed a technique that supports users to perform freehand sketches in real time by presenting blurred images in the background [Lee et al. 2011]. They first divide the sketch into patches. The length and direction of the lines in the patch are expressed using a binary coherent edge (BiCE) descriptor. The sketch is compared with a large number of images in the database, based on this descriptor. As a result, the image that has a certain number of similar patches is extracted. Thereafter, in order to generate a blurred image to be displayed in the background for guiding the user, the extracted images are synthesized according to the score of the similarity. Since Lee et al.'s work [Lee et al. 2011] aims to provide some hints to users, only local and rough shape matching was considered; therefore, it cannot be directly used for the retrieval of clothes. However, for future work, their technique can be employed for supporting the user to draw a sketch for retrieval in our system.

(a) Sketch drawn by Subject 1. (b) Best matching image of (a) searched by Subject 1.

(c) Sketch drawn by Subject 1. (d) Best matching image of (c) searched by Subject 1.

(e) Sketch drawn by Subject 3. (f) Best matching image of (e) searched by Subject 3.

Figure 2: *Results of preliminary experiments.*

3 Primary Study: Features for Retrieval

Research on image retrieval based on sketches has received a lot of attention in recent years. Various feature descriptors aiming at more accurate retrieval have been proposed [Hu et al. 2010; Chen et al. 2009; Eitz et al. 2010]. In sketch-based image retrieval, it is important that the employed descriptors can deal with the concepts represented in users' sketches. The descriptors should also capture well the features to be extracted from the images. The conceptual match is a key to realize high-quality, sketch-based image retrieval.

We have conducted a primary study to observe what is represented in a user's sketch. Three female university students with varying drawing skills participated in the study. They were asked to freely draw sketches of the skirts they wanted to buy. After drawing the sketches, they were instructed to manually retrieve the best matching skirts from websites. Sketch drawing and image retrieval were fixed in the above order, and we did not recommend any specific websites but asked the subjects to do their best to search for the skirt images that best matched their sketches.

Figure 2 shows three examples of the sketches drawn by the three subjects, respectively. Figure 2 (b), (d), and (f) are the best matching images searched by the subjects themselves for their sketches in Figure 2 (a), (c), and (e), respectively. Note that the best matching skirt is slightly different from what the subject had imagined when she drew the sketch. In Figure 2 (b), the orientation of the texture in the sketch is different from that of the best matching image. At the post-task interview, the subject answered that she could not find one with exactly the same texture, but this one was the most similar one she could find.

By observing the sketches and conducting the post-task interview, we have found that the typical features the subjects wanted to present in their sketches are the shape, texture, and one-point decoration. The shape includes the outline, flare, pleats, or other details. Subject 1 (Figure 2 (a)) has indicated the checked texture by crisscrossed lines. A ribbon on the waist on the upper right corner (Fig. 2 (c) and (d)), which is called a one-point decoration, adds an accent to the design of the skirt. We employ these features for matching between a sketch and a skirt image. In the following sections, we design and implement an algorithm for extracting these features from both the image and the sketch.

4 Design of Feature Vector for Retrieval

4.1 Shape

We employ the Fourier descriptor to represent the shape feature of a skirt. Although there are several choices, such as the shape context mentioned in Section 2, we choose the Fourier descriptor [Zhang and Lu 2001] to take advantage of controllability over the level of detail. The feature match between sketch and image is done in the frequency domain. The coarse and fine contours of the skirts are well described as low- and high-frequency components. The low-frequency components possess the information on the overall shape of a skirt, and the high-frequency components contain the information on the folds, pleats, and flare of a skirt. Both of them are vital for retrieving a particular skirt.

Conventional edge detection filters, such as the Laplacian filter and Canny filter, do not extract the contours directly. They produce edges only. The contour is a subset of the edges representing the boundary between an object and its background. We first extract the region of a skirt and sample a sequence of points on the contour. The frequency components of these contour points are finally calculated by the Fourier transform. Since the skirt images and sketches usually contain much "noise," the processes are not straightforward in practical terms. The detailed processes are described in the following subsections.

4.2 Robust Skirt Region Extraction

The Canny edge filter often produces unnecessary edges for extracting the contour. When the skirt region has a texture, the unnecessary internal edges of the skirt are also extracted. When the texture consists of the same color as that of the background, then the extracted contour might not be continuous. Since object region segmentation from a cluttered background is a challenging problem itself, we assume that skirt images have simple non-textured backgrounds in this work. In order to obtain the contour, we first extract the foreground from the background region with the following steps:

1. Apply the Canny edge filter to a skirt image to obtain the edges.
2. Apply the morphological closing (erosion of dilation) operation to the edge images obtained in Step 1.
3. Fill all the closed contours to obtain connected components.
4. Select the largest connected component as the skirt region.

When performing the morphological closing, the times of iterative dilation (erosion) are important. Since we only need the largest skirt region and do not care about how the edges inside the skirt region are connected, we can simply choose a sufficient large number as the times iteration experimentally. In the current implementation, three times of iteration have already produced the ideal results.

4.3 Fast Fourier Transform

Chain-code tracking converts the contour into a sequence of the contour points. To apply the fast Fourier transform (FFT) to the contour points, the number of sample points is restricted as factorial of 2 for implementation. The points are sampled from the original sequence by the equal arc-length sampling algorithm. A sampled point $(x(t), y(t))$ with the index t of 1 to 2^n is converted to $r(t)$, as follows:

$$r(t) = ([x(t) - x_c]^2 + [y(t) - y_c]^2)^{1/2} \tag{1}$$

where x_c and y_c comprise the gravity point of all the contour points. The frequency components of the contour are calculated by applying 1D FFT. For accelerating the retrieval at runtime, the Fourier descriptors are computed for all the skirt images in the database at the preprocessing stage and stored with the images. The distance between a skirt image and a sketch is defined as the L_2 distance of the Fourier descriptor, with frequency components as its dimension.

4.4 Texture

We adopt Gabor filters for generating the texture feature descriptor. Gabor filters are known to be useful for detecting the frequency and direction of a texture. As shown in Figure 4, by using Gabor filters of varying frequencies and orientations, we should be able to identify a texture. The Gabor filters used can be described as:

$$\begin{aligned} & g(x, y; \lambda, \theta, \psi, \sigma, \gamma) \\ & = exp(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}) cos(2\pi \frac{x'}{\lambda} + \psi) \\ & x' = x cos\theta + y sin\theta \\ & y' = -x sin\theta + y cos\theta \end{aligned} \tag{2}$$

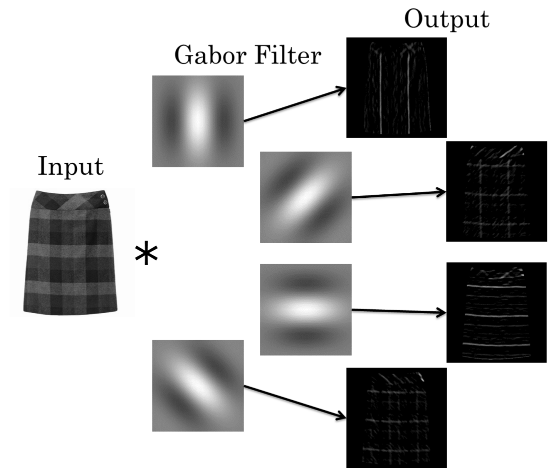


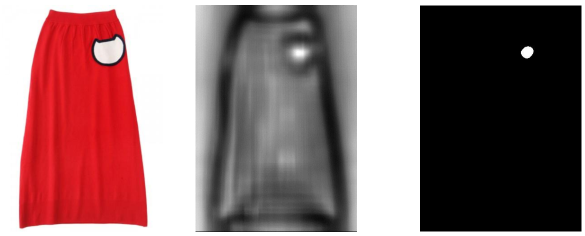
Figure 3: *Gabor filtering.*

where θ indicates the angle and $1/\lambda$ the frequency. If a texture's orientation and frequency match those of a particular Gabor filter, then a high power remains in the filtered image. We use 20 Gabor filters, with 5 successive frequency levels and 4 different directions. For each of the 20 Gabor filters, two window sizes are used. The mean and standard deviation of the filtered image are computed, and an 80 $(2 \times 4 \times 5 \times 2)$ dimensional feature vector is obtained. The vector represents the remaining power after being filtered by Gabor filters, which characterizes the texture of a skirt in an image and a sketch.

4.5 One Point Decoration

We employ the saliency map to capture the one-point decoration feature (Figure 5). The saliency map indicates the degree of how salient a region is, compared to its surroundings in an image. When a region has different characteristics of color, brightness, or texture from its surroundings, the region has a high value in the saliency map. Because the one-point decoration region should have a feature different from those of its surroundings, it should indicate a high value in the saliency map.

We first generate a saliency map of the sketch and images in the database. Next, we cut out the region indicating a high value in the saliency map, and then detect the position of the one-point decoration. The feature vector of the one-point decoration consists of its distance and angle from the gravity center of the contours and horizontal line, respectively. We design the feature vector in this way, based on the consideration that the position of the decoration is more important than the detailed features of the decoration itself. During the post-task interview in the primary study, according to the subjects, what is important is whether there is a one-point decoration or not and where it is; they do not care much about its detailed features. For example, either a ribbon or a small flower would not make a big difference for them.



(a) Input image (b) Saliency map (c) Extracted one-point decoration region

Figure 4: *One-point decoration feature.*

For the calculation of the saliency map, we use Klein et al.'s [Klein and Frintrop 2011] method, which provides good maps with respect to speed and performance. The saliency map might be affected by background noise, which is an undesirable result. Therefore, we use the masked image resulting from contour extraction (in Section 4.1.1) to separate the background from the skirt region first, and then only the skirt regions are considered for generating the saliency map. The one-point decoration is the region that stands out with a large value in the saliency map.

5 Experiments

5.1 Implementation of Interactive Retrieval System

In order to verify how our designed feature vectors contribute to the retrieval of a skirt, we developed the prototype of an interactive sketch, skirt retrieval system. We used a desktop PC (OS: Windows 7(x64), MM: 4 GB) with a tablet display monitor (WACOM DTU-710) and implemented with Visual Studio 2010 and OpenCV 2.4.0.

Previous to the experiments, we built a skirt image database of 247 skirt images that were manually collected. Since automatic segmentation in an image is a difficult problem itself, we collected skirt images with plain backgrounds only. Our proposed feature vectors were calculated when an image was stored in the database. Each feature, shape, texture, or one-point decoration had its own unit and range of values. They were normalized to 0 to 1 with respect to the range of the feature values.

Figure 6 is a snapshot of the operating window of the system. On the top is the menu for specifying the style and attributes of the sketch, as well as for adjusting the parameters for retrieval. On the right is the output window for displaying the retrieved results, and on the left is the canvas for the user to draw the sketches. When a user sketches a skirt on the canvas, the system immediately calculates our proposed feature vector from the sketch. The calculated feature vector is compared with those in the database. The approximate nearest neighbor (ANN) searching algorithm [Arya et al. 1998] is used for accelerating the retrieval. With the current database size, the time required for retrieving the top five candidates is within several seconds. The weighting coefficients for combining the features of shape, texture, and one-point decoration are user controllable. The weighting values can be set to 0 for a particular feature, which means that the subject does not mind this feature. For example, if a user wants a skirt of a particular shape, but any texture is fine, and prefers to have a one-point decoration, then she can achieve the result by setting the weight value of the texture to 0. The top N images in the order of similarity are displayed on the result window. In the current system, the default value of N is set to 5, and the user is allowed to change it interactively.

5.2 Results

Examples of the experimental results are shown in Figures 7 – 10. In Figure 7, a short skirt with a pleated and checked texture was drawn by the user. All the top matched images except for the third one satisfied the subject's intention. The similarity in shape of the third image was much higher than those of the others, but the similarity in texture was lower than those of the others, which was why the image was selected as the third similar one. In the example shown in Figure 8, a long skirt with three sections and a border texture was sketched. Obviously, the first and the fourth skirts matched the sketch the best in both shape and texture. Figure 9 is the result when the weight of the shape is doubled. We can see the two most similar skirts come to the top. After analyzing the statistics of the features for all the skirts in the database, we have found that the deviation of the shape feature is twice as large as that of the texture feature. Therefore, in case we want to balance the shape and texture, it is necessary to raise the weight value of the shape feature to be twice that of the texture. This accounts for the fact that we could achieve the ideal results shown in Figure 9. We will improve the weight computing method by taking into consideration the deviation of each feature. Figure 10 shows an example with a one-point decoration. The skirt with a one-point decoration at a similar position was searched as the top matched one.

We also tried shirt image retrieval. We wanted to test if our designed feature vector is applicable to other kinds of clothes. Since long-sleeved shirts are widely bought on fashion sites and have relatively complex shapes, we chose shirts as the target. We collected 33 shirt images (long sleeves: 20, short sleeves: 6, and three-quarter length sleeves: 7) for the database. The sketch of the shirt is shown in Figure 11. We employed the features of shape, texture, and one-point decoration in the retrieval for shirts in the same way as in the case of skirts. The results look reasonable. The third shirt was selected for the similarity of shape, and the fifth one was chosen due to the similarity of texture.

6 Conclusions and Future Work

We have designed a feature vector for skirt image retrieval from sketches. The observation of the user sketches in our primary study has indicated the shape, texture, and one-point decoration intended in the sketches. The implemented prototype system has presented intuitively similar images with the input sketches. We have also confirmed that the feature vector has the possibility for application to other types of clothes.

For future work, we need to refine the feature vector and its weightings. The current texture feature cannot accurately distinguish between polka dots and checked textures. Garments other than skirts will require new features to be described. The weighting is also critical for better image retrieval. The weighting should depend heavily on the sensitivity of the users. By modeling the sensitivity for each feature, more intuitive matching images will be given by our system.

We have also tried to use shirt images in the experiment as the first extension to other clothing. We plan to observe how users draw shirts in their sketches. There would be both common and uncommon attributes with those of skirts. The design of the feature vector would be different, depending on the types of clothes.

In addition, we need to employ more advanced segmentation

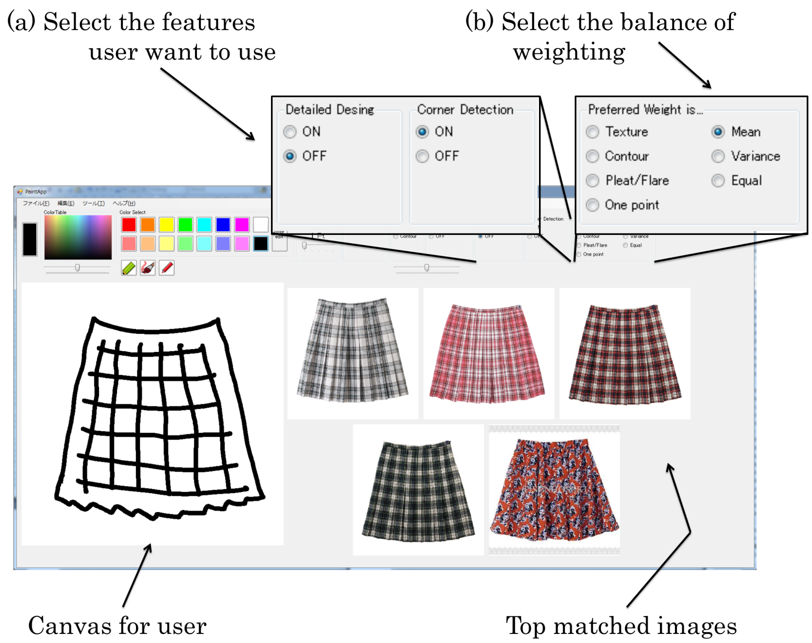


Figure 5: *Sketch interface.*

(a) Sketch input (b) System output (the top rank from the left)

Figure 6: *Retrieval result 1.*

(a) Sketch input (b) System output (the top rank from the left)

Figure 7: *Retrieval result 2.*

(a) Sketch input (b) System output (the top rank from the left)

Figure 8: *Retrieval result 3.*

method for extracting skirt regions from cluttered backgrounds. In this work, the database consists of images with simple backgrounds only. If we can deal with images with arbitrary backgrounds, then larger database can be built by collecting cloth images automatically.

Acknowledgements

This work was partially supported by KAKENHI 25280037.

(a) Sketch input (b) System output (the top rank from the left)

Figure 9: *Retrieval result 4.*

(a) Sketch input (b) System output (the top rank from the left)

Figure 10: *Retrieval result 5.*

References

ARYA, S., MOUNT, D., SILVERMAN, R., AND WU, A. 1998. An optimal algorithm for approximate nearest neighbor searching in fixed dimensions. *Journal of the ACM 45*, 6, 891–923.

CHEN, T., CHENG, M.-M., TAN, P., SHAMIR, A., AND HU, S.-M. 2009. Sketch2photo: Internet image montage. *ACM Transactions on Graphics (TOG) 28*, 5. Article 124.

EITZ, M., HILDEBRAND, K., BOUBEKEUR, T., AND ALEXA, M. 2010. Sketch based image retrieval: Benchmark and bag-of-features descriptors. *IEEE Transactions on Visualization and Computer Graphics (TVCG) 17*, 11, 1624–1636.

HU, R., BARNARD, M., AND COLLOMOSSE, J. P. 2010. Gradient field descriptor for sketch based retrieval and localization. In *IEEE International Conference on Image Processing (ICIP)*, 1025–1028.

KLEIN, D. A., AND FRINTROP, S. 2011. Center-surround divergence of feature statistics for salient object detection. In *International Conference on Computer Vision (ICCV)*, 2214–2219.

LEE, Y. J., ZITNICK, C. L., AND COHEN, M. F. 2011. Shadowdraw: real-time user guidance for freehand drawing. *ACM Transactions on Graphics (TOG) 30*, 4. Article 27.

LIU, S., SONG, Z., LIU, G., XU, C., LU, H., AND YAN, S. 2012. Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3330–3337.

TSENG, C. H., HUNG, S. S., AND TSAY, J. J. 2009. An efficient garment visual search based on shape context. *WSEAS Transactions on Computers 8*, 7, 1195–1204.

YAMAGUCHI, K., KIAPOUR, M. H., ORTIZ, L. E., AND BERG, T. L. 2012. Parsing clothing in fashion photographs. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3570–3577.

ZHANG, D., AND LU, G. 2001. A comparison of shape retrieval using fourier descriptors and short-time fourier descriptors. In *IEEE Advances in Multimedia Information Processing*, 855–860.