Proceedings of the International Conference on Machine Vision and Machine Learning
Prague, Czech Republic, August 14-15, 2014
Paper No. 126

Generating Fashion Design Sketches from Images

Jie Xu, P.Y. Mok*, R.W.Y. Yee
The Institute of Textiles and Clothing, the Hong Kong Polytechnic University,
Hunghom, Hong Kong
jarek.xu@connect.polyu.hk; tracy.mok@polyu.edu.hk; rachel.yee@polyu.edu.hk

Abstract **–** Taking an image processing perspective, a framework is proposed in this paper to generate fashion sketches, in the form of vector graphs, from images. The framework starts from an image segmentation step that separates the target products from the clutter background. It follows with an edge detection step, in which a new method is developed by integrating results of two methods for better feature preserving and reducing false responses. A series of interactive tools are developed to edit defects in the result maps, for example, the defect due to occlusions. Finally, an algorithm is proposed to convert the result maps into vector graphs. The proposed framework is not confined to fashion sketch generation, but is applicable to other image edge detection applications. It is particularly effective for line extractions or the images affected by illuminations.

Keywords**:** Fashion sketches, Image processing, Edge detection, Image segmentation.

1. Introduction

In the digital age, customers are always connected and they can easily have access to information via multiple channels multiple devices. The ability to reaching, addressing the needs, and engaging today's empowered customers is essential for any businesses in a fast changing marketplace. Accelerated new product development is thus crucial for survival. In the fashion industry, the transformation is profound. What used to be a norm of three- to six-month production cycle – the time it takes to design, manufacture, and distribute clothing to stores – has collapsed to two to three weeks. The phenomenon is popularly known as 'fast fashion' that a design idea just recently shown on runway in Paris or Milan can be made available on racks in high street stores, in a wide range of sizes, at affordable prices within the space of a month.

To shorten the product development cycle, some have researched on intelligent computer aided design (CAD). For example, Mok et al. (2013) proposed a CAD system that non-professional users can use to create new designs by selecting designs with features appear interesting to them. Another example is KnitSketch proposed by Ma et al. (2011), which mimics a paper-and-pencil-like sketching method to improve design efficiency and accuracy. These prototypes assist designers creating new designs, in the form of sketches, from the very beginning. It is interesting to note that another way to speed up the design process is by turning an existing product image into sketch. In practice, designers often get inspirations from existing products; the job of a designer can be described as creatively integrating existing product silhouette with new details or incorporating interesting product details to a contemporary shape/look. In this paper, taking the perspective of image processing, a framework is proposed to generate product sketches from images. Fashion products are selected to illustrate the research idea; the algorithms developed are not limited to fashion items, but applicable to other product categories.

The idea of converting images into editable vector graphs echoes with the concept of edge detection in image processing. In ideal cases, edge detection algorithms reduce the data of any input images to the minimum by preserving only edges indicating the boundaries of objects. However, edge detection is not a trivial task. Due to the diversity of input image conditions, most edge detectors suffer from the problems fragmented edges, missing edge and false detection (Lindeberg, 1998). We propose a pipeline in this paper by integrating two edge detection approaches for better preserving important features and converting such into fashion sketches.

Traditionally, edges are detected according to the intensity gradient magnitudes, such as Sobel (1970) and Prewitt (1970) operators. These operators perform poorly when edges were blurred and noisy, because the operators with simple structures cannot provide immunities to noises (Basu, 2002). More sophisticated operators were proposed later. Marr and Hildreth (1980) introduced Gaussian filter into edge detection, and the Gaussian filter can suppress noise to a certain degree. However, Marr and Hildreth (1980) adopted zero crossings in the second order derivative of intensities to find edges, but not all zero-crossings correspond to edges, so false edges still resulted. Canny (1986) also introduced Gaussian filter into his algorithm, different from Marr and Hildreth, he used first order derivative of intensities to detect edges and adopted hysteresis thresholding to enhance the weak edges detection ability. Canny operator has been widely used until now. But it still suffers from some limitations. The main problem of Canny operator and other gradient-based operators is that it operates only based on the intensities of the image, which is different from human's visual system (HVS). Therefore, the results derived from gradient-based operators are usually a bit different from what people expected. For example, gradient-based operators are dependent on illumination conditions. Accidental edges, such as shadow and shading edges, may be caused by profound illumination effects (Gevers, 2004). For another example, gradient-based operators often cause double responses for line features, such as the seam line in a skirt.

Compared with approaches based on intensity gradients, another approach for edge detection is phase congruency-based, detecting edges in the frequency domain. Morrone and Burr (1988) pointed out the relationship between the features and the phase of the Fourier components of the image intensity, i.e. the phase congruent points corresponding to the features in the image. They also indicated the corresponding features are agreed with what perceived by HVS. However, the phase congruent is hard to calculate, so the local energy whose response is proportional to phase congruent can be employed to instead the phase congruent (Robins, 1999). Robbins and Owen (1997) indicated the local energy is usually calculated by convolving an image with a quadrature pair of filters in the spatial domain. Nevertheless, the local energy is only suitable for one dimension signal processing. For 2-dimension problems, such as in images, the total energy can be used to detect features. The total energy can be usually approximated as sum of local energies at several orientations, and the local energy with orientations is called Oriented Energy (OE). The OE can achieve relative better results and solve the problems like double responses for line features in gradient-based edge detectors. However, because the OE is approximated by limited orientations, the edge which contains parts in different orientations may have different response magnitudes. If the threshold is not appropriate, it could still lead to discontinuous results.

2. Methodology: Image to Sketch Processing Algorithms

In this paper, we proposed a pipeline that converts input images into editable sketches by integrating the result from Canny operator and that from OE-based edge detection method. Fig. 1 shows the workflow of the proposed framework.

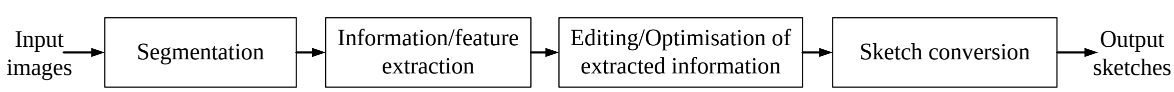


Fig.1. Workflow of the proposed framework.

The framework starts from an image segmentation step that separates the target products from the clutter background. It follows with an edge detection step, in which a new edge-detection method is developed by integrating results of two methods for better feature preserving and reducing false responses. A series of interactive tools are developed to edit defects in the result maps, for example, the defect due to occlusions. Finally, an algorithm is proposed to convert the result maps into vector graphs.

2.1 Target Segmentation

Fashion images such as street shots and runway photos, target fashion products are usually with a clutter background. It thus is necessary to separate target products from the background to facilitate latter processing. In the field of image segmentation, many methods have been proposed over the years. However, there is no completely automatic method in image segmentation (Varshney et al., 2009). Therefore, semi-automatic segmentation technique is adopted. In the area of interactive image segmentation, some powerful algorithms have been developed, such as Rother et al. (2004) and Boykov and Jolly (2001). In this project, the target product is extracted by a tool developed based on the method of Vezhnevets and Konouchine (2005). By user selecting a small number of seed pixels on the images as foreground and background, the tool labels the rest pixels automatically. As shown in Fig. 2(a), the red points are seed pixels in the foreground and the blue points are seed pixels in the background. For the selection of seeds, the tool allows (but not requires) users to dynamically adjust the seeds according to displayed segmentation results. Fig. 2(b) shows the boundary between the foreground and background. Fig. 2(c) is the extracted image which can be used for subsequent processing.

(a) (b) (c)

Fig.2. The process of target segmentation.

2.2 Information/Feature Extraction

Target products are composed of two parts: boundaries and inner features. The boundaries of the target products are extracted from the segmentation step. Next, the inner features of the products are extracted. In order to obtain continued edges, a method is proposed to fusion the result maps from the OE-based method and that from Canny operator, as show in Fig. 3.

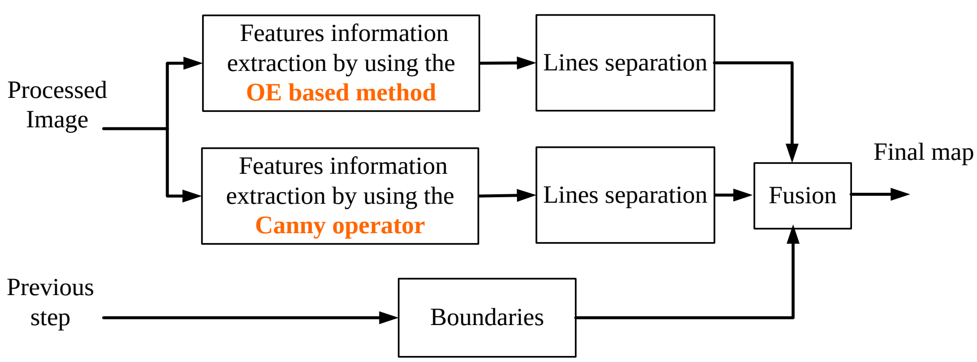
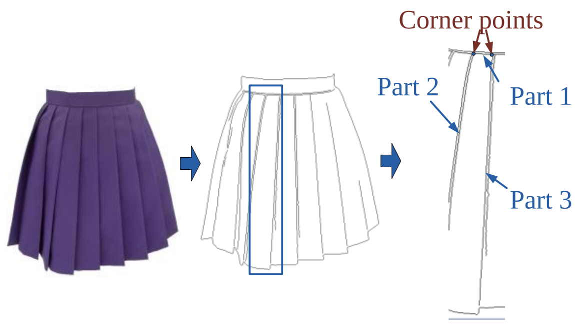


Fig.3. Information/feature extraction.

In the fusion process, the result maps from Canny operator can be analogy as a *dictionary*; each edge (line) in the result maps of the OE-based method will look up corresponding ones from the dictionary. If more complete lines are available, they are replaced with the corresponding ones in the dictionary. In the looking up process, the integrity of the corresponding lines in the two result maps is compared. In the result maps, only lines with a series of continued pixels points are considered. A result line may contain two or more parts which are originated from different features or false edges, especially in the results from

Canny operator, as shown in Fig. 4(c). This affects the integrity comparisons of lines and may introduce false edges to the results. In calculating OE-based results, Gabor filter is usually used to make convolutions with the image, because of the multi-scale and multi-orientation properties in Gabor filter. Moreover, in order to enhance feature visibility, Difference-of-Gaussian (DoG) filter is used to pre-process the given image.



(a) (b) (c)

Fig.4. (a) is the input image (b) is the result map from Canny operator (c) is an enlarged view of the selected region of (b).

The steps of information/feature extraction by OE-based method are as follows:

1) Input a processed image I (an example is as shown in Fig. 5(a)) and define a DoG filter D, so $I_D = I * D$, where '*' is the convolution operator. The DoG filter D is defined as a difference of two Gaussian functions as

$$D(x,y,\sigma)=\frac{1}{2\pi\sigma^2}\exp^{-(x^2+y^2)/(2\sigma^2)}-\frac{1}{2\pi K\sigma^2}\exp^{-(x^2+y^2)/(2K\sigma^2)} \quad (1)$$

where (x, y) is the pixel location. σ represents the standard deviation in the Gaussian function. K is the coefficient between two Gaussian functions. K =2 is used in this paper.

2) Define a filter F, which is combined by a quadrature pair of Gabor filters, as

$$F = G_{\theta,0,\sigma,\lambda,\gamma}(x,y) + iG_{\theta,\frac{\pi}{2},\sigma,\lambda,\gamma}(x,y) \quad (2)$$

where θ, φ, σ, λ and γ represent the orientation, the phase offset, the sigma/standard deviation of the Gaussian envelop, the wave length and the spatial aspect ratio, respectively.

$$G_{\theta,\varphi,\sigma,\lambda,\gamma}(x,y)=\exp(-\frac{x'^2+\gamma^2y'^2}{2\sigma^2})\cos(2\pi\frac{x'}{\lambda}+\varphi) \quad (3)$$

where $x' = x\cos\theta + y\sin\theta$ and $y' = -x\sin\theta + y\cos\theta$.

3) $J = I_D * F$ =($I_D * G_{\theta,0,\sigma,\lambda,\gamma}(x,y)$)+i ($I_D * G_{\theta,\frac{\pi}{2},\sigma,\lambda,\gamma}(x,y)$) and the oriented energy

4)

$$E_\theta = |J| = \sqrt{(I_D * G_{\theta,0,\sigma,\lambda,\gamma}(x,y))^2 + (I_D * G_{\theta,\frac{\pi}{2},\sigma,\lambda,\gamma}(x,y))^2} \quad (4)$$

In this paper, 12 directions was adopted, so $\theta = 0, \pi / 6, \pi / 3, \dots , 11\pi /6$;

5) The total energy $E = \sum E_\theta$,$\theta = 0, \pi / 6, \pi / 3, \dots , 11\pi / 6$, and the energy magnitude map *Mag*

can be obtained (as shown in Fig. 5(b)).

6) Make non-maximum suppression and get a map with 'sharp' candidate edges Mag_s. Then normalize $Mag_{s.}$ (as shown in Fig. 5 (c)).
7) In the Mag_s, the magnitude of the boundaries is very strong. However, boundaries do not need to extract and the features in inner region should be focused on. Therefore, a mask M (as shown in Fig. 5 (e)) obtained from the segmentation step is used to get an energy magnitude map Mag_p which only contains data of energy magnitudes of the inner region.
8) Construct an array Ar, and let Ar contains the values larger than zero in Mag_p.
9) Split all values in Ar into two classes using the method of Otsu (1979) as background and foreground, and get the threshold T.
10) The hysteresis thresholding used in the Canny operator is adopted. In this paper, the high threshold $T_{high} = T$ (defined in step 8), and the low threshold T_{low} is still defined by Otsu's method, the scope is the background class defined is step 8. Therefore, the edges map (as shown in Fig. 5(g)) can be obtained from Mag_s.
11) In the edges map, lines that are shorter than one-fiftieth of the sum of the image width and height are excluded. It reduces noises and accelerates subsequent fusion process.

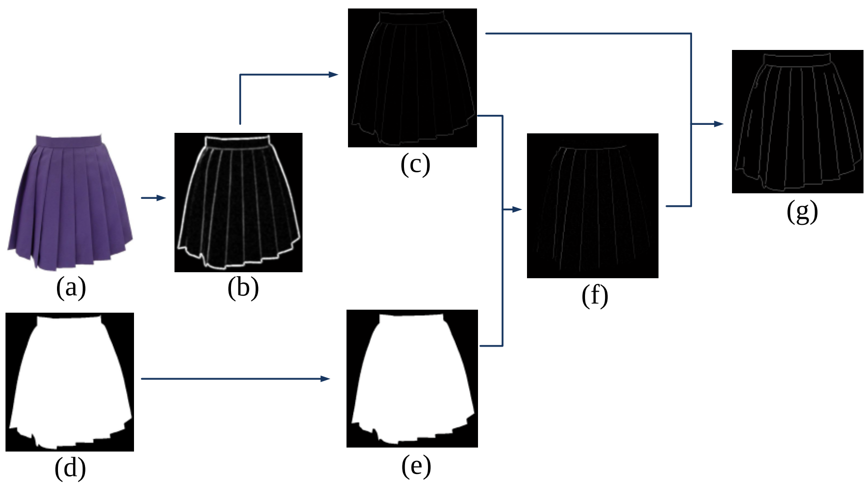


Fig. 5. The process of the OE-based information/feature extraction.

Canny operator mainly contains four steps including *smoothing*, *find gradients*, *non-maximum suppression* and *hysteresis thresholding*. In this paper, Canny operator is modified as *step 6* of the above process. The gradient magnitudes obtained from the *non-maximum suppression* step will be used to make dot products with shrunken mask. Meanwhile, Otsu method is also used provide the reference threshold value. The high threshold equals to quarter of the reference threshold value, and the low threshold is defined as twenty percent of the high threshold in this paper.

Corner points are used to decompose a result line (as shown in Fig. 4(c)), which is considered as a combination by a series of continued pixels points in the result map. We adopted He and Yung's (2008) algorithm for the corner point detection. First, the absolute curvature of each point on the result lines is computed. Then, the local maxima of the absolute curvature are chosen as initial corner candidates. Among these candidates, round corners are removed using adaptive curvature threshold. Finally, false corners caused by quantization noise and trivial details are eliminated. After all corner points are calculated, they are converted as start points or end points of new lines; each line including all pixel points is then stored.

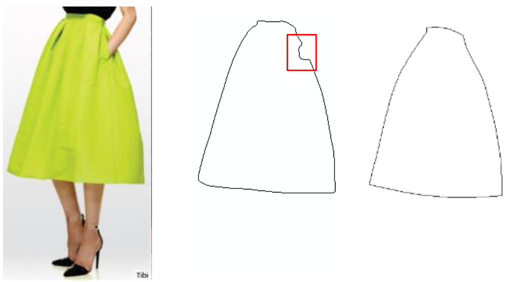
In the fusion process, the result maps from OE-based method and Canny operator are fused firstly. The line group G_{CO} from Canny operator is considered as reference dictionary, the line group G_{OE} from OE-based method are looked up from the dictionary. If the corresponding lines in the dictionary are longer than the ones in result maps of OE-based method, they are replaced the corresponding ones in the dictionary. In order to reduce effects of location errors caused from edge detection process, the lines in G_{OE} will be dilated a little. The fusion algorithm is listed below.

```
Define a null group G_f
For each line L_OE in G_OE
    Calculate pixel number of L_OE, denoted as N_OE.
    Find all lines in G_CO longer than L_OE, and construct a group named G_temp.
    If  G_temp is not null
        Obtain DL_OE by dilating L_OE 1 pixel.
        For each line L_temp in G_temp
            Calculate matched pixels number N_m between DL_OE and L_temp.
            If N_m / N_OE > 0.7, L_temp is selected into the candidates group G_cand
            If G_cand is null, L_OE is put into G_f
            else The longest line in G_cand is chosen into G_f
    else  L_OE is put into G_f
```

After constructing G_f, the boundary from segmentation step is used to replace the boundary in G_f.

2.3 Editing/Optimisation of Extracted Information

A fused result map is obtained from the information/feature extraction step. All result lines in the map with separated corner points are recorded in a database. The extraction results are usually affected by some other factors of the images, such as illumination and shadow, thus a global threshold for the whole image may not be ideal for all features in different places. It is possible to extract excess information (false edges) in some places but miss information in other places. Therefore, a simple interactive tool is developed to allow users select and delete excess information. For the missed edge, the information/ feature algorithm is modified and applied to user selected region. Fragmentation, discontinued lines, is another defect of edge detection. Although the proposed Canny and OE-based integrated edge detection method provide more complete lines, it cannot eliminate fragmentation in some cases. To solve the problem, a simple algorithm is developed to connect user selected segments automatically. In this algorithm, all start/end points of the selected segments are computed, the neighbour points in different segments are then connected by spline curves. Besides, the defect caused by partial occlusions of the input images is common. For runway photos, fashion products are often occluded by some body parts, such as hands (see an example in Fig. 6(a)). These occlusions cause errors in the resulting fashion product silhouettes (as shown in Fig. 6(b)). An interactive tool is developed that users can select the defect region (e.g. red rectangle in Fig. 6(b)), then the defect area is replace with a spline, which is tangent to other parts of the silhouette (as shown in Fig. 6(c)).



(a) (b) (c)

Fig.6. An example of the defect area editing.

2.4. Sketch Conversion

The result maps and revised information are recorded as decomposed lines. An algorithm is developed to compose these lines as spline curves and output results as vector graphs with Drawing Exchange Format (DXF).

3. Results

In this section, some examples obtained from the proposed system are shown in Fig.7. The images in the 2nd row in Fig.7 are target fashion products extracted from input images in the 1st row. The results in the 3rd row are obtained by the proposed method and the thresholds are calculated automatically by Otsu's method. Based on the above results, users can interactively lower or increase the threshold for better results (as shown in Fig.7 (4a) and (4e)). The results in the 5th row are the final results which have been edited by users. For example, excess information was removed in Fig.7 (5a) and a covered area has been modified in Fig.7 (5c) and (5f). All final results are vector graphs and they can be further processed by most CAD software which support DXF.

Fig.7. Results.

Because of the involvement of some interactive manipulations, it is hard to evaluate objectively the speed of the method. Table 1 gives response times for the part of information/feature extraction for the images shown in Fig.7. It is shown that 2-3 seconds are needed to extract information/features from fashion image inputs.

Table1. The response times for images in Fig.7

Image No.	Fig.7(a)	Fig.7(b)	Fig.7(c)	Fig.7(d)	Fig.7(e)	Fig.7(f)
Resolution (pixels)	150×350	330×580	200×400	270×460	240×580	330×600
Response times (s)	0.86	1.70	3.89	2.44	3.11	2.05
Configuration: Intel Core2 Duo CPU E8400 3GHz, Ram 4G						

4. Conclusions

In this paper, we have presented a framework to generate sketches of fashion products from images. The key new development is a new edge detection method, which combines advantages of both OE-based edge detector and Canny operator. It has shown the proposed framework is effective to generate fashion sketches from images of different conditions. The method is not only applicable to fashion products but other products. It has demonstrated effective edge extraction, especially when precise line extraction is needed or when input images are affected by illuminations.

Acknowledgements

The work described in this paper was financially support by The Hong Kong Polytechnic University (Project code: RT7B, A-PL78).

References

Basu M. (2002). Gaussian-based edge-detection methods-a survey. IEEE Transactions on Systems, Man, and Cybernetics-part C, 32(3), 252–26

Boykov Y. and Jolly M. (2001). Interactive graph cuts for optimal boundary & region segmentation of objects in ND images. ICCV 2001. Vol. 1. 105–112.

Canny J. (1986). A computational app roach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8 (6) : 679-698.

Gevers T. (2004). Robust Segmentation and Tracking of Colored Objects in Video. IEEE Transactions on Circuits and Systems for Video Technology, 14(6), 776–781.

He X.C. and Yung H.C. (2008). Corner detector based on global and local curvature properties. Optical Engineering, 47(5), 057008.

Lindeberg T.(1998). Edge detection and ridge detection with automatic scale selection. International Journal of Computer Vision, 30(2), 117-154.

Ma C.X., Liu Y.J., Yang H.Y., Teng D.X., Wang X.A. and Dai G.Z. (2011). KnitSketch: A Sketch Pad for Conceptual Design of 2D Garment Patterns, IEEE Transactions on Automation Science and Engineering. 8(2), pp 431-437

Marr D. and Hildreth E. (1980). Theory of edge detection. Proc. Royal Society of London, B, 207, pp. 187–217.

Morrone M. and Burr D. (1988). “Feature Detection in Human Vision: A Phase Dependent Energy Model,” Proc. Royal Soc. of London B, vol. 235, pp. 221-245.

Mok P.Y., Xu J., Wang X.X., Fan J.T., Kwok Y.L. and Xin John. (2013). An IGA-based design support system for realistic and practical fashion designs, Computer-Aided Design. 45, 1442–1458.

Otsu N. (1979). "A threshold selection method from gray-level histograms". IEEE Trans. Sys., Man., Cyber. 9 (1): 62–66.

Prewitt J.M.S. (1970). Object Enhancement and Extraction, in Picture Processing and Psychopictorics, Lipkin B.S., and Rosenfeld, A.(eds.),Academic Press, New York.

Robbins B. and Owens, R. (1997). 2D feature detection via local energy. Image and Vision Computing, 15(5), 353–368.

Robins M.J. (1999). Local energy feature tracing in digit al Images and Volumes. PhD thesis, Univ. of Western Australia, pp.5- 21.

Rother C., Kolmogorov V. and Blake A. (2004). “Grabcut: Interactive foreground extraction using iterated graph cuts,” Proc.ACM.Siggraph.

Sobel I.E. (1970). Camera Models and Machine Perception, Ph.D dissertation, Stanford University, Palo Alto, California.

Varshney S.S., Rajpal N. and Ravindar P. (2009) Comparative study of image segmentation techniques and object matching using segmentation, in International Conference on Methods and Models in Computer Science, pp. 1–6.

Vezhnevets V. and Konouchine V. (2005). GrowCut: Interactive multi-label ND image segmentation by cellular automata. proc. of Graphicon.