An Efficient Garment Visual Search based on Shape Context

¹Chin-Hsien Tseng, ²Shao-Shin Hung, ³Jyh-Jong Tsay and ⁴Derchian Tsaih

 ^{1,3}Department of Computer Science and Information Engineering National Chung Cheng University, Chiayi, Taiwan 621
²Department of Computer Science and Information Engineering, WuFeng Institute of Technology, Chiayi, Taiwan, R.O.C
⁴Department of Electronic Commerce Management, Nanhua University, Chiayi, Taiwan, R.O.C {¹tsengch, ²hss, ³tsay}@cs.ccu.edu.tw; ⁴dtsaih @mail.nhu.edu.tw

Abstract: - In recent years, the theoretical models of mass consumer behavior have change to buy from websites rather than in stores. Because the high-growing of e-commerce, a new demand emerges: the special-purpose search engine for searching goods from network shop. How to meet the customer's requirement in product search is an import problem. Although it is easy for human eyes to determine the existence of clothes styles, recognizing it automatically from a computer program is not a trivial problem. Our work focuses on the garment retrieval from the e-shopping database, which supports feature-based retrieval by shape categories and styles. Traditionally the rigid shape-based algorithms unable to apply well on garment images. Because the clothing is essentially a non-rigid soft object: it is apt to self-occlusion, folding, and has deformation among every part (such as sleeve and tube). While producing deformation, it also influenced by light which lead to various kinds of shade at clothes, and the surface might include various kinds of pattern, texture, little piece, and decorate, these will all cause the great interference on image analysis.

Key-Words: Visual Search, Visual Similarity, Garment, CBIR, Non-Rigid Matching, Shape Context

1 Introduction

Advances in telecommunications and computer technologies in recent years have made computer networks an integral part of the economic E-commerce provides infrastructure. multiple benefits to the consumers in form of availability of goods at lower cost, wider choice and saves time. Many shopping search engines are designed to check prices at various online stores or locate e-commerce outlets by category, like Google's shopping engine: Froogle (http://www.google.com/products), Yahoo (http://hk.shopping.yahoo.com/), Shopping and BizRate (http://www.bizrate.com/).

Besides, the world wide web has been witnessed through the last few years as a platform for a plethora of aggregated information. This information extends way beyond its original hypertextual form including, most commonly, streaming audio, video, dynamic web content, etc. Notably, web is also manifesting its presence in ubiquitous computing devices, extending thereby its user community. Many shopping search engines are designed to check prices at various online stores or locate e-commerce outlets by category, like Google's shopping engine.

On the other side, with the prevalence of digital imaging devices such as webcams, phone cameras

and digital cameras, image data are now explosively increased. An emerging issue is how to browse and retrieve this daunting volume of images. The search engine based on text query performs well to such goods with concrete name as the electrons, books, and etc. However, consumers are difficult to describe their demands clearly for the products with numerous styles and have no concrete descriptions (e.g. dress, furniture). Users can only input the brief and rough keywords to query about, and often spend a lot of time, select one by one in the tediously goods retrieval list. It is clear that if users do not know the search keyword or image unable to correlated to text labels, or reach the wrong text label, the text query will be useless. When the demand is indeterminate, the products searching perhaps unsuitable in the way like choose checkbox or input products attribute value; the course is unable to be so rational. Obviously the pictures of clothing provide more important cues then descript text in products retrieval. Traditional text-based querying is not adequate to deal with the garment retrieval because text features are less accurate and might takes mismatch between the user's expression and the

user's understanding and expectation. Against to above problems, our research aims to design an image-based query system which allows user to upload simple clothes snapshot or paints the sketch with hands then it uses the image to find alike images in a picture collection. In traditional image retrieval system, image content might refer to low-level features such as shapes, colors, and textures. But in the garment search, consumers primarily take care of the style of clothing, and then just consider the color and texture. Therefore we mainly use the shape feature for image matching, not use any color or texture information.

In fact, clothing choice is an important way in which people communicate their individuality, identity, tastes, status, age, wealth, *et al.* [31, 41-43]. Many sociologists consider clothing to be a form of communication, but one that uses a code of "low semanticity." Direct interpretation of the meaning of fashion is difficult because the semantics are inherently ambiguous and because the meaning varies among individuals, times and situations [6]. Nevertheless, what you wear certainly says *something* and machine perception should at least be able to detect coarse levels of information.

In applications related to consumer image collections [44, 45], clothing color features have been characterized by the correlogram of the colors in a rectangular region surrounding a detected face. For assisted tagging of all faces in the collection, combining face with body features provides a 3-5% improvement over using just body features. However, segmenting the clothing region continues to be a challenge; all of the methods above simply extract clothing features from a box located beneath the face, although Song and Leung [44] adjust the box position based on other recognized faces and attempt to exclude flesh.

In this paper, we explore the use of computer vision to recognize classes and attributes of clothing, specifically shirts, for a variety of applications including identifying a store customer's taste and spending profile and recommending "similar" or "different" clothes that match his or her fashion preferences. The recommendation application could be instantiated on a variety of platforms—e.g., as a web-based application, or a real-time assistant in a fitting room display [30, 32].

The key components of the system are the clothes classification and clothes similarity measurement algorithms. Clothes recognition is difficult in a number of ways: (1) the social nature of the problem definition. (2) The real-time requirement of the algorithm. (3) The complexity of the vision problems involved in clothes recognition, for example, the high intra-class variation and deformable configurations of the clothes. These points will be addressed in detail in the following section.

The application described appears at first glance to resemble a typical computer vision problem. Several interesting problems arise, however, in the course of implementing such a system, an example of a class of problems which we term *clothe vision problems*. As opposed to traditional problems in computer vision, social vision problems require us to appreciate and understand how humans perceive objects that have social meaning.

The following discussion highlights the underlying problems in the proposed application. First, given an image of a shirt, what features need to be recognized? A computer vision algorithm may be able to accurately detect the distance between a shirt's buttons, but do people care about this feature when looking at shirts? Thus, we must ensure that our system recognize the features that people care about. We refer to this as the clothe-driven feature selection problem. Secondly, once the relevant features are known, how well can the vision system detect those features in a piece of clothing? This falls in the traditional computer vision space, but clothing provides some unique challenges due to its large intra-class diversity. This is in addition to the fact that both human bodies and clothes are deformable and algorithm must be invariant to these the deformations.

Finally, once we have examined the relevant features of a piece of clothing, how do we determine which of our stored images are similar? Is a blue dress-shirt more similar to a green dress-shirt or a blue t-shirt? We refer to this as the *clothe-driven feature comparison problem*.

For a set of articles of clothing with good light conditions and uniform background (no illuminations and background clutter), we want to query by an image or by a hand-drawing picture and retrieval a set of images ranked by the shape similarity of the query image. The shape similarity of clothes is defined by their category and style of components (e.g. neckline, sleeve, cuffs, and inner patterns).

Although humans can easily detect style of clothes and estimate component parts of clothing, these works are inherently difficult for a computer.

There are three major challenges in cloth representations (as shown in Figure 1).

(1) Geometric deformations: clothes do not have static form and are very flexible.

(2) Photometric variability: large variety of colors, shading effects, and textures.

(3) Topological configurations: a combinatorial number of cloth designs – T-shirts, jackets, pockets, zips, suits, sweaters, coats where cloth components may be reconfigured and combined to yield new styles.

Our primary contribution in this work is a strategy for solving garment matching on non-rigid geometric deformations like folding, wrinkle, self-occlusion, non-affine distortion, and independent of photometric variability.



Fig 1. Problems in clothes shape matching

Traditionally the techniques of Search by Image are called by CBIR (Content-based Image Retrieval). The searching subject of CBIR was mainly similar to rigid images or shapes. When the deformation of searching target restricted on a small range, it already exists a lot of algorithms result in a well retrieval. But those algorithms unable to apply well on garment images, because the clothing is a non-rigid soft object in essence: it is apt to self-occlusion, folding, and has deformation among every part (such as sleeve and tube). While producing deformation, it also often influenced by light which lead to various kinds of shade at clothes, and the surface might include various kinds of pattern, texture, little piece, and decorate, these will all cause the great interference on image analysis. In general, the rigid objects detection method is inappropriate for handling this problem. For example, Lowe's SIFT [16] is one of the most robust registration method of correspondence matching. In our experiment it performs well in small distortion (Figure 2-a) but has no match when occurs non-rigid geometric deformations like foldings and wrinkles (Figure 2-b).

2.1 Clothes Recognition

Clothes recognition is to judge how similar two pieces of clothes images are and therefore to indicate how likely they are from the same individual. There are three major steps for clothes recognition: clothes detection and segmentation, clothes representation (or feature extraction), and similarity computation based on extracted features.

Generally speaking, there are three types of information. context The first type is appearance-based, such as a person's hair style or the clothes he is wearing; the second type is logic-based, for instance, different faces in one picture belong to different persons or some people are more likely to be pictured together; the third type the meta-data for pictures such as the is picture-taken-time. This context information is often bv human observers consciously used or unconsciously. It is very temping to investigate hoe to build algorithms which can utilize this context information effectively to improve human recognition accuracy.

Being able to identify people is important for automatic organizing and retrieving photo albums and for security applications, where face recognition has been playing a major role. But reliable face recognition is still a challenging problem after many research efforts. Recent works on identifying people consider context information besides faces, such as clothes, picture taken-time and some logical constraints.

As the features are concerned, there are generally two types of features being extracted: local features and global features. Most local features are selected based on some kind of local extrema (e.g. with maximum entropy or maximum change), which cannot work if the clothes under consideration is a smooth colored region without textures or patterns. The global feature like color histogram suffers when lighting changes Clothes are often folded and therefore create false edges and self-shadows, which create difficulties for orientation histograms. Thus some more effective features are desired. Literature [1] solved this problem by using code-word (representative patches) frequency as feature vectors. Overlapped small image patches (e.g. 7×7 pixel patches with two neighboring patches 3 pixels apart) are taken from each normalized clothes piece.

2 Related Works



Fig 2. Problems of rigid objects detection method.

In order to get rid of noise and make the computation efficient, principle component analysis (PCA) is used to reduce the dimensionality of these vectors. Each small patch is represented by projections under the first k principle components. Vector quantization (e.g. K-means clustering) is then run on these N k-dimensional vectors to obtain code-words.

Zhang et al. [2] present a clothing recognition system that augments clothes recommendation and fashion exploration using the intelligent multi-view vision technology. they observed that the skin tone from a person's face is usually similar to the skin tone of arms. Therefore first ran an efficient face detector to detect the location of the person's face to estimate the rough arm area for sleeve length recognition. Then explore features by the number of Harris corner points.

2.2 Clothes Modeling

Agn'es Borr'as et al. [4] present a two-stage process for providing an automatic description of clothing according to the color, texture and structural composition of garments. In segmentation process, they discriminate the texture zones by MPEG-7 texture descriptor and apply a pixel-based technique to split plain regions of the same colour. In interpretation process, they describe a clothing composition by a model structured as an attributed graph where the nodes represent the garment regions and the edges represent their relationship. The similarity between graphs are computed by node costs (area, texture presence and colour homogeneity) and edge costs (spatial position and include relation).

Chen et al. [5] present a context sensitive grammar in an And-Or graph representation which will produce a large set of composite graphical templates to account for the wide variabilities of cloth configurations, such as T-shirts, jackets, etc. They ask an artist to draw sketches on a set of dressed people, and decompose the sketches into categories of components. Each component has a number of distinct sub-templates that serve as leaf nodes in a big And-Or graph. An And-node represents a decomposition of the graph into sub-configurations with Markov relations for context and constraints. An Or-node is a switch for choosing one out of a set of alternative And-nodes.

3.3 Shape Context

Belongie et al. [8, 9, 10] have introduced the shape context descriptor, which characterizes a particular point location on the shape contour by using statistics of other contour points seen by this point in quantized angular and distance intervals. Corresponding points on two different shapes have a similar relative position in each shape. The obtained view of a single point p forms a 2D histogram called the shape context of p.

The original shape context shape descriptor is sensitive to image distortions. Liu and Chen [11] make an attempt to diffuse, or spread the contribution of each contour point. This process softens the boundary of the bins and is essentially equivalent to applying a low-pass filter to the original shape context. Mori et al. [16] extended the shape context descriptor by encoding more descriptive information than point counts in the histogram bins. In this work, each contour point is attached a unit length tangent vector that is the direction of the edge. Thayananthan et al. [13] included edge orientation information with the shape context descriptor. The edge orientation is the gradient vector of the edge pixels around a shape context point. They also proposed the continuity constraint in the correspondence estimation by using Viterbi algorithm. According to the experiments of Schwer [17], the Viterbi point matching algorithm seems to be a poor fit for the shape context matching problem. Some other approach to enforce the continuity constraint is needed. Carneiro et al. [14] proposed a variation of shape context by combining it with local image features (the multi-scale phase based [15] and SIFT [16] features). A vote in a specific histogram bin is weighted by a function that decreases with distance



Fig 3. Process of feature extraction.

3 Methodology

3.1 Overview

Our solution on clothes retrieval consists of the following phases (see in Figure 3): Phase 1 is *Segmentation*, then phase 2 is *Feature extraction* and the final phase is *Shape matching*.

We firstly apply a segmentation approach that partly eliminates the influences of textures, wrinkle, shading, decorations and inner patterns. Note that the segmentation process might bring some side-effect like lost information of hems and sewing lines. After that, the original image is transferred into segmentation line graph and can be view as a sketchy and coarse representation. In the feature extraction and shape matching steps, we treat the image as the 2D point set and try to compute shape correspondences and similarity.

Our feature extraction and matching methods are based on the Shape Context descriptor. The process can be divided into the following steps: 1) Sampling points, 2) Shape Context, 3) Bipartite Graph Matching. The detail was discussed in Section 4.3 and 4.4. For a query image, we calculate its shape context descriptor and compare to all images in the database. Therefore, we retrieval an image list ranking by the shape similarity to the query image.

3.2 Segmentation

In order to get good features from the closing and eliminate the partly interferences from textures, wrinkle, shading, decorations and inner patterns, we first perform the image segmentation. Vincent Martin has experiment [20] on six recent and powerful segmentation algorithms: Meanshift, J Segmentation (JSeg), Efficient Graph-Based Image Segmentation (EGBIS), Color Structure Code (CSC), Statistical Merging Segmentation (SRM), and Region Growing. He tests each algorithm on each image of the Berkeley Segmentation Dataset [21]. In this database, 6000 hand-labeled segmentations of 500 color Corel images from 30 human subjects have been collected. According to his test results of optimization stage on each training image for each algorithm, the JSeg approach performs well contour continuity and texture clustering then others on garment images (An example result is shown in Fig 14). Therefore, we picked JSeg, which used unsupervised color and texture segmentation to segment clothes images.

As shown in Fig 4, the segmentation groups and simplifies inner patterns, and is effective in eliminating textures, wrinkle, and shading. The drawback of this method is obviously: the hems and sewing lines are also eliminated. Since the intensity of hems and sewing lines are comparatively weak in the whole image, they will be discarded after texture grouping processing. Examples of bad segmentations such as fragments and over growing are shown in Fig 5.

3.3 Feature Extraction

(1) Sampling points

The shape context analysis begins by taking N samples from the edge elements on the shape. These points can be on internal or external contours. In the original article of Belongie et al. deemed that the point set need not and typically will not, correspond to key points such as maxima of curvature or inflection points.



Fig 4. (a) original image (b) Canny-Deriche edge detection (c) Jseg segmentation.



Fig 5. Example of bad segmentations.



Fig 6. (a) After Harris corner detection (b) Sample from remain points.

Although they prefer that the samples be roughly uniform in spacing, but think this is also not critical. In our experiment on clothes image analysis, we find when use a small amount of sampling points; it is helpful to promote the retrieval accurate rate by lying sample points uniformed on the edge line. In order to get a "good" layout we first do the Harris corner detection, sort the finding corner points, and then get the best n corners under a minimum distance gap. If the number of corners less then the demanded sampling number, we then delete the points around the corner points and random select remain points up to the demanded number. A sampling instance is shown in Fig 6, after that an object shape can be represented by a discrete set P.

(2) Shape Context

Consider the *n*-1 vectors originate from a point to all other sample points on this shape. These vectors express the configuration of the entire shape relation to the reference point. To achieve rotation invariance, we first calculate the most frequent angle in polar space and treat the tangent vector related to the angle as the positive x-axis. For each points we rotate their position to the reference frame. To achieve scale invariance, all radial distances are normalized by the mean distance α , which is obtained by averaging norms of *n*-1 vectors.

$$\alpha = \sum_{k=1}^{n} \left\| p_k - p_i \right\| / n$$

For a point p_i on the shape, we compute a coarse histogram h_i of the relative coordinates of the remaining n-1 points. This histogram is defined to be the shape context of p_i . Where *n* is the number of sample points and *k* is equal to 60 (5 log bins multiply 12 angle bins).

Here uses bins that are uniform in log-polar space, making the descriptor more sensitive to positions of nearby sample points than to those of points farther away. In order to soften the boundary of the bins a low-pass filter is applied to the shape context histogram. The cost of matching two points is calculated by χ^2 test statistic, then the correspondence problem can be solved by minimizing the total cost of matching.

4. Experimental Results

4.1 Clothes Dataset

Since there are no existing benchmarks for closing image evaluation, we manually collect images of all kinds from the clothing e-shopping website to measure the retrieval performance. Our preliminary clothing dataset contains 90 images which are classified into 13 categories according to their style. For each category we designate a hand-drawing sketch as the classified template as shown in Fig 5.

4.2 Recall Comparison

The recall performance metric is defined as the proportion of the relevant material actually retrieved. Our retrieval performance was evaluated by using the simplifized recall comparison called *First Tier* (FT) and *Second Tier* (ST). For a class with *C* members, K = C - I for the first tier, and K = 2*(C - I) for the second tier. The first tier statistic indicates the recall for the smallest K that could possibly include 100% of the models in the query class, while the second tier is a little less stringent.

$$FT = \frac{R_1}{C - 1} \times 100\%$$
$$ST = \frac{R_2}{2(C - 1)} \times 100\%$$

4.3 Comparison with Different Sampling Points

The result of performance comparison is shown in Table 1. We first experiment by random selecting 200

sampling points, for each image in the database the average matching time is 2.4 minutes in C+++ program and total take 3.6 hours on 90 images. The second experiment use 100 uniform spread sampling points and the average matching time for each image reduces to 1/4, moreover, the matching result in FT and ST all raise to 0.7%. In template matching, the uniform spread sampling takes about 1/6 average matching time then random selection and gets the same matching rate.



Fig 5. Garment templates for classification.

Table 2 shows the template matching retrieval rate in 12 clothes categories. The matching accuracy is all the same in each categories, it might due to the template shape is too simple but it also shows the uniform sampling really bring improve. Also, Table 3 shows retrieval matching rate of 12 clothes categories in FT and ST. Shapes with outstanding appearances such as pants and hoods have the higher distinguishing rate.

5. Conclusion

In this thesis, we explore the shape context techniques and discussed how to apply it for better search results in clothes image retrieval. Most computation is performed in pre-processing stage, and the resulting feature description is encoded to point cloud. Each feature includes 100 coordinates (using 200 Bytes). Our retrieval performance achieves 32.73% retrieval rate in average. Each models take about 2.4 minutes for retrieving on a PC with a Pentium IV 3.0 GHz CPU and GeForce 6600 video card. The algorithms used in our thesis are efficient and simple, therefore this method is applicable to cooperate with the other approaches to get higher performance.

Sampling points	100 (uniform spread)	200 (random selection)
Template average	4 sec (C++)	23 sec (C++)
time (image		
number $= 15$)		
Template matching	0.366667	0.366667
	(33/90)	(33/90)
Full matching	34 sec (C++)	2.4 min (C++)
average time	126 sec	9.5min
(image number =	(Matlab)	(Matlab)
90)		
Full matching First	0.44	0.37
Tier		
Full matching	0.51	0.44
Second Tier		

Table 1. Retrieval performance.

Point number	100	200	
Skirt (15)	7	6	
Shirt (3)	0	0	
Pants (5)	4	4	
Shorts (8)	3	3	
Blazers (3)	1	1	
Hoodies (4)	4	4	
Vest (4)	0	0	
Sleeveless (3)	0	0	
Dresses (5)	4	4	
U-Neck (3)	2	2	
V-Neck (4)	2	2	
T-shirt (19)	6	7	
Others (14)	-	-	
Match number (90)	33	33	
Matching Rate	0.366667	0.366667	

Table 2. Template matching retrieval rate.

Point number	100 (uniform		200 (random	
	spread)		selection)	
Tier	First	Second	First	Second
Skirt (15)	0.35	0.51	0.28	0.44
Shirt (3)	0.33	0.33	0.33	0.33
Pants (5)	0.64	0.8	0.44	0.6
Shorts (8)	0.28	0.39	0.30	0.36
Blazers (3)	0.55	0.56	0.55	0.55
Hoodies (4)	0.63	0.69	0.63	0.75
Vest (4)	0.55	0.56	0.33	0.33
Sleeveless (3)	0.33	0.33	0.33	0.33
Dresses (5)	0.44	0.60	0.36	0.52
U-Neck (3)	0.44	0.44	0.33	0.33
V-Neck (4)	0.38	0.38	0.25	0.25
T-shirt (19)	0.38	0.58	0.32	0.54
Others (14)	-	-	-	-
Total Average	0.44	0.51	0.37	0.44
Rate				

Table 3. Full matching retrieval rate.



Fig 7. Results of image retrieval: Left side is the query image and right side lists retrieval images ranked by the similarity.

In the future, we would like to extend this method for higher retrieval rates and lower storages spaces. There are three directions for further work: First, an applicable preprocess for precedent segmentation will help analyzer further to get more accurate results. By Integrating Canny edge with segmentation boundary the shape descriptor will be more distinguishable for similar and dissimilar models. Second, in order to refine the retrieval results, we can cooperate with many feature detection techniques, e.g., Local Self-Similarities [22], Incomplete Contour Representations [23], and Minimal Shape Prototypes [24]. Finally, in order to refine the search result and improve accuracy, a relevant feedback technique can be applied to. Relevant feedback uses Support Vector Machine (SVM) algorithm to adjust the weights of similarity evaluation metric. According to user's desire and feedback, the relevant results are brought closer and the irrelevant results are pushed farther. Exploring ways combine relevant feedback techniques in our system is another interesting work.



Fig 8. Template Matching Result

References:

- [1] D. Anguelov, K. Lee, S. B. Gokturk, B. Sumengen. Contextual identity recognition in personal photo albums, *Proc. CVPR*, 2007.
- [2] Y. Song, T. Leung. Context-aided human recognition clustering, *Proc. ECCV*, 2006.
- [3] Wei Zhang, Bo Begole, Maurice Chu, Juan Liu, Nick Yee, Real-time clothes comparison based on multi-view vision, *the 2nd ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC)*, 2008.
- [4] Borràs Agnès, Tous Francesc, Lladós Josep, Vanrell Maria, High-level clothes description based on color-texture and structural features, *1st. Iberian Conference on Pattern Recognition and Image Analysis IbPRIA* 2003.
- [5] Hong Chen, Zi Jian Xu, Zi Qiang Liu, Song Chun Zhu, Composite templates for cloth modeling and sketching, *Proc. CVPR*, 2006.
- [6] Tak-Kan Lau, Irwin King, An Image Database for the Fashion, Textile, and Clothing Industry in Hong Kong, *Proceedings of the Third Asian Conference on Computer Vision-Volume I*, , January 08-10, 1998, pp.410-417.
- [7] Chiou-Ting Hsu and Ming-Chou Shih, Content-Based Image Retrieval by Interest Points Matching and Geometric Hashing, *Proc. Of SPIE Photonics Asia Conference*, vol. 4925, Shanghai, China, Oct. 2002, pp. 80-90,.
- [8] S. Belongie and J. Malik, Matching with Shape Contexts, *IEEE proc. Content-Based Access of Image and Video Libraries*, 2000, pp. 20–26.
- [9] G. Mori, S. Belongie and J. Malik, Shape Contexts Enable Efficient Retrieval of Similar Shapes, *Proc. IEEE Comp. Soc. Conf.*

Computer Vision and Pattern Recognition, vol. 1, 2001, pp. 723-730.

- [10] Serge Belongie, Jitendra Malik, Jan Puzicha, Shape matching and object recognition using shape contexts, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002.
- [11] D. Liu and T. Chen, Soft Shape Context for Iterative Closest Point Registration, *International Conf. Image Processing*, vol. 2, 2004, pp. 24–27.
- [12] G. Mori, S. Belongie and J. Malik, Efficient Shape Matching Using Shape Contexts, *IEEE trans. Pattern Analysis and Machine Intelligence*, vol. 27, 2005, pp. 1832–1837.
- [13] A. Thayananthan, B. Stenger, Shape Context and Chamfer Matching in Cluttered Scenes, *Proc. IEEE Comp. Soc. Conf. Computer Vision and Pattern Recognition*, vol. 1, 2003, pp. 127–133.
- [14] G. Carneiro and A. D. Jepson, Pruning Local Feature Correspondences Using Shape Context, *Proc. Int. Conf. Pattern Recognition*, vol. 3, 2004, pp. 16–19.
- [15] G. Carneiro and A. Jepson. Multi-scale phase-based local features, *Proc.CVPR*, Madison, Wisconsin, USA, 2003.
- [16] Lowe D. G., Object recognition from local scale-invariant features, *Proceedings of International Conference on Computer Vision*, 1999.
- [17] Schwer. P., Shape context in clutter, University of California San Diego, CSE Graduate Student, Project Report.
- [18] Canny J., A computational approach to edge detection, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, 1986, pp.679-714.
- [19] Lowe D. G., Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, vol. 60, no. 2, 2004, pp. 91-110.
- [20] Image segmentation experimentations: http://www-sop.inria.fr/pulsar/personnel/Vince nt.Martin/segmentation/
- [21] Berkeley Segmentation Datase: http://www.eecs.berkeley.edu/Research/Projec ts/CS/vision/grouping/segbench/
- [22] Eli Shechtman and Michal Irani, Matching local self-similarities across images and videos, *Proc. CVPR*, The Weizmann Institute of Science, 2007.
- [23] Anarta Ghosh and Nicolai Petkov, Robustness of shape descriptors to incomplete contour.

representations, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005.

- [24] Thuresson J. and Carlsson S., Finding object categories in cluttered. images using minimal shape prototypes, *Proceedings of the 13th Scandinavian Conference*, SCIA 2003.
- [25] Xiao-Dong Zhuang and Nikos E. Mastorakis, A Magneto-Statics Inspired transform for structure representation and analysis of digital images, WSEAS Trans. Computers, issue 5, vol. 8, 2009, pp. 874-883.
- [26] Ran Jin, Zhuojun Dong, Face Recognition based on Multi-scale Singular Value Features, WSEAS Trans. Computers, issue 5, vol. 8, 2009, pp. 865-873.
- [27] X.D. Zhuang, and N.E. Mastorakis, The curling vector field transform of gray-scale images: a magneto-static inspired approach, *WSEAS Trans. Computers*, issue 3, vol. 7, 2008 pp. 147-15.
- [28] Liang Zhou, Yu Sun, and Jianguo Zheng, Automated color image edge detection using improved PCNN model, *WSEAS Trans. Computers*, issue 4, vol. 7, 2008, pp. 184-189.
- [29] F. Han and S.C. Zhu, A two-level generative model for cloth representation and shape-from-shading, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.29, no.7, July 2007, pp 1230-1243.
- [30] W. Zhang, T. Matsumoto, J. Liu, M. Chu., B. Begole, An intelligent fitting room using multi-camera perception, IUI, 2008.
- [31] Barnard, M. *Fashion as Communication*, 2nd ed., Routledge, 2002.
- [32] Davis, F. Fashion, Culture and Identity, University Of Chicago Press, 1994.
- [33] S. Ioffe and D. A. Forsyth, "Probabilistic methods for finding people", *IJCV*, vol.43, no. 1, 2001, pp. 45-68.
- [34] R. Ronfard, C. Schmid and B. Triggs, Learning to parse pictures of people, *ECCV*, 2002.
- [35] M.W. Lee and I. Cohen, Human upper body pose estimation in static images, *ECCV*, 2004.
- [36] H. Chen, Z.Q. Liu, C. Rose, Y.Q. Xu, H.Y. Shum and D. Salesin, Example-based composite sketching of human portraits, *Proc. 3rd Int'l Symp. NPAR*, 2004, pp.95-153.
- [37] L. Zhao, Dressed Human Modeling, Detection, and Parts Localization", *Ph.D Dissertation*, CMU, July, 2001.
- [38] H. N. Ng et al, Computer graphics techniques for modeling cloth, *IEEE Computer Graphics* & *Applications*, vol.16, 1996, pp.28-42.
- [39] R. X. Y. Tian, W. Liu, F. Wen, and X. Tang. A face annotation framework with partial

clustering and interactive labeling. *Proc. CVPR*, 2007.

- [40] C.E. Guo, S.C. Zhu, and Y.N. Wu, A mathematical theory for primal sketch and sketchability, *Proc. ICCV*, 2003.
- [41] K. Nicolaides. *The Natural Way to Draw*, Houghton Mifflin Co. Boston, 1941.
- [42] K. Bhat, C. D. Twigg, J. K. Hodgins, P. K. Khosla, Z. Popvic, and S. M. Seitz., Estimating cloth simulation parameters from video, *Proc. Symposium on Computer Animation*, 2003.
- [43] R. X. Y. Tian, W. Liu, F. Wen, and X. Tang., A face annotation framework with partial clustering and interactive labeling, *Proc. CVPR*, 2007.
- [44] Y. Song and T. Leung. Context-aided human recognition- clustering. *Proc. ECCV*, 2006.
- [45] N. Sprague and J. Luo. Clothed people detection in still images, *Proc. ICPR*, 2002.