Efficient Cloth Pattern Recognition Using Random Ferns

Inseong HWANG†, Seungwoo JEON†, Beobkeun CHO†, Nonmembers, and Yoonsik CHOE†, Member

SUMMARY This paper proposes a novel image classification scheme for cloth pattern recognition. The rotation and scale invariant delta-HOG (DHOG)-based descriptor and the entire recognition process using random ferns with this descriptor are proposed independent from pose and scale changes. These methods consider maximum orientation and various radii of a circular patch window for fast and efficient classification even when cloth patches are rotated and the scale is changed. It exhibits good performance in cloth pattern recognition experiments. It found a greater number of similar cloth patches than dense-SIFT in 20 tests out of a total of 36 query tests. In addition, the proposed method is much faster than dense-SIFT in both training and testing; its time consumption is decreased by 57.7% in training and 41.4% in testing. The proposed method, therefore, is expected to contribute to real-time cloth searching service applications that update vast numbers of cloth images posted on the Internet.

key words: circular patch window, cloth pattern recognition, delta-HOG, DHOG, rotation and scale invariant random ferns

1. Introduction

While the smart mobile environment is in rapid development, the mobile shopping industry is growing every day. Visual search technologies for cloth are particularly getting more attention. Reducing both the errors of the existing search technology and computational costs for object recognition are key issues for enhancing the object recognition performance which is still insufficient for actual cloth searches [1]–[5]. For example, the standard approach, e.g., SIFT [6], is to build affine-invariant descriptors of local image patches and to compare them across images. This usually involves fine scale selection, rotation correction, and intensity normalization. It results in high computational overhead without satisfactory performance and often requires handcrafting the descriptors to achieve insensitivity to specific kinds of distortion.

On the other hand, Ozuysal et al. introduced random ferns [7], [8], which is used for learning class-conditional distribution, \(P(F|C_i)\), and its training time is very fast; it only grows linearly with fern size. In many object detection problems, the run-time performance is very important, therefore random ferns are more popular. In addition, it is very simple to implement, and it performs as well as SIFT. The cloth pattern recognition method using this random ferns, however, has to consider slightly wrinkled cloth patches because these patches are from the middle area of real cloth images which can be a little wrinkled. This is because we use an HOG [9] based descriptor. As it is known, the HOG is a kind of template method that is independent of pose and location changes and is not sensitive to lighting changes. Therefore, using this HOG and employing invariance to scaling and rotation, a rotation invariant HOG-based descriptor and delta-HOG (DHOG) are proposed for fast and efficient classification even when the cloth patches are rotated and rescaled; the rotation and scale invariant random ferns method is also proposed using this descriptor, which is fast and efficient classification method.

2. Proposed Method

In this paper, the entire recognition process with rotation and scale invariant is shown. The FAST-9 algorithm [10] is used for key-feature point detection.

2.1 Rotation and Scale Invariant Delta-HOG-Based Descriptor for Cloth Pattern Recognition

The DHOG construction process is shown in Fig. 1. A circular patch window is first proposed for the descriptor because it can prevent pixel data changes in the patch window whenever the window is rotated for detecting the pose changes of object. To speed up the calculation of polar coordinates in the circular patch window, every pixel location and its set of polar parameters(radius and angle from the window center) in the circular patch window is stored in a table in PC memory; therefore, we can simply read it through relative addressing using the table without calculating it.

Image gradients are computed using Prewitt operators over the input image, \(I(x, y)\), by

\[
G_y(x, y) = I(x, y) \ast P_y, \quad G_x(x, y) = I(x, y) \ast P_x, \quad (1)
\]

\(G_y(x, y)\) and \(G_x(x, y)\) are the gradient in the \(y\) and \(x\) directions, respectively. The gradient value is normalizing by

\[
\text{Intensity normalization} = \frac{G_y(x, y)}{\sqrt{G_x(x, y)^2 + G_y(x, y)^2}}
\]

\(G_y(x, y)\) and \(G_x(x, y)\) are normalized intensity values.

On the other hand, the gradient amplitude \(G = \sqrt{G_x(x, y)^2 + G_y(x, y)^2}\) is mapped to the angle \(\theta\) using

\[
\theta = \arctan\left(\frac{G_y(x, y)}{G_x(x, y)}\right)
\]

As a result, \(\theta\) is in the range of \([0, 2\pi]\) radians.

\(\theta\) is normalized and processed with a quantization step size of 10 degrees for rotation, and \(\theta\) is divided into \(\frac{360}{18}\) segments. The histogram is accumulated in each segment, so the final histogram is accumulated in a total of 18 bins. The \(\theta\) is divided into 18 bins, which is divided into 360 degrees/18 = 20 degrees. The final histogram is accumulated in a total of 18 bins.

The proposed method is fast and efficient, especially for cloth pattern recognition because the cloth images are usually wrinkled. The proposed method is expected to contribute to real-time cloth searching service applications.
where \( G_x \) and \( G_y \) are the gradient image maps, and \( P_x \) and \( P_y \) are the Prewitt operators. From the gradient maps, the orientation, \( \phi \), is calculated at the Cartesian coordinates in the image region, \( B \):

\[
\phi(x, y) = \tan^{-1} \frac{G_y(x, y)}{G_x(x, y)}, \quad \forall (x, y) \in B \tag{2}
\]

Subsequently, each gradient casts a vote for one orientation bin according to its orientation, \( \phi(x, y) \). For the orientation case, the orientation, \( \phi(x, y) \), is discretized into \( m \) orientation bins which, in this example, is 8. In addition, each gradient casts a vote for its orientation bin using its magnitude, \( \text{mag}(x, y) \), as weight:

\[
\text{mag}(x, y) = ||G_x(x, y), G_y(x, y)||_2, \tag{3}
\]

An orientation histogram is formed from the gradient orientations of sample points within the region. Inspired by SIFT descriptor, a rotation invariant HOG-based local binary feature (RIHLBF) vector using the maximum orientation of the local image region is proposed. As can be seen in Fig. 1, the maximum orientation (MO) in the orientation histogram corresponds to the dominant directions of the local gradients and the circular window is turned to the MO degree for detecting objects in a specific category that may appear in images under different rotations.

Subsequently, each gradient casts a vote again for one spatial and orientation bin according to its location, \((x, y)\), inside region \( B \) and its orientation, \( \phi(x, y) \); for this case, the image region is divided into a \( 2 \times 2 \) sub-region (a, b, c, d). Therefore, each gradient casts a vote for its sub-region and orientation bin using its magnitude, \( \text{mag}(x, y) \), and orientation, \( \phi(x, y) \), as weight:

\[
\text{HOGB}(s, b) = \sum \text{mag}(x, y) \quad \forall (x, y) \in B, \tag{4}
\]

where \( \text{HOGB} \) is the HOG-based descriptor computed in the image region, \( B \), and \( b \) and \( s \) are the orientation bin and sub-region alphabet. The result is a histogram of 32 elements.

On the other hand, a simple split function for random ferns is originally obtained from the difference between any two orientation bin locations:

\[
f(x) = \begin{cases} 
1 & \text{HOGB}(s_i, b_i) > \text{HOGB}(s_j, b_j) \\
0 & \text{HOGB}(s_i, b_i) \leq \text{HOGB}(s_j, b_j)
\end{cases} \tag{5}
\]

When considering the MO, it is possible to precisely predict where each pairwise bin in an original HOG should appear in the transformed image. Therefore \( f(x) \) has to be transformed to \( f_{\text{RIH}} \) using the MO:

\[
f_{\text{RIH}}(x) = T_{\text{MO}}(f(x)), \tag{6}
\]

Let the shifted \( b' \) be \((b_1 - MO + 2\pi)\text{mod}2\pi\) and \( b'' \) be \((b_2 - MO + 2\pi)\text{mod}2\pi\),

\[
f_{\text{RIH}}(x) = \begin{cases} 
1 & \text{HOGB}(s'_i, b'_i) > \text{HOGB}(s'_j, b'_j) \\
0 & \text{HOGB}(s'_i, b'_i) \leq \text{HOGB}(s'_j, b'_j)
\end{cases} \tag{7}
\]

On the other hand, identifying textured patches surrounding key-points across images acquired under widely varying scale conditions is the key-point of many computer vision algorithms. To acquire these scale-varied correspondences, the circular bounding window is resized with various scales in the range of a \([0.5, 1.5]\)-ratio, because the classifier is often involved in comparing objects that have been resized. To handle scale variations for \( \text{HOGB}(s, b) \), a circular window with a different scale and radius \( R \) is applied to the keypoints. In other words, the scale space is analyzed by changing the scanning window size rather than iteratively reducing the image size, as in the pyramid structure shown in Fig. 2.

The number of pixels in the circular bounding window is changed according to the different scale and it is preserved as a different profile. In this paper, the standard radius of the circular bounding window is 16 pixels, which is considered to have no scale change. The following layers are obtained by gradually changing the radius, \( R \), according to the scale range. We generate all possible shifts of an initial circular window with the \([8, 24]\)-pixel range of \( R \) which represents a \([0.5, 1.5]\)-ratio. Note that as we do not have to downsample the image, no aliasing occurs. However, the calculation of a lot of HOGs according to every circular window with radius \( R \) for each keypoint requires a high computational cost; this creates a burden for real-time applications. Therefore, the multi-scale layered DHOG descriptor is proposed for resolving this computational problem, as shown in Fig. 2.

In Fig. 2, the relationship between the DHOG and the HOG is explained. Let the orientation be \( \psi \); \( \text{HOGB} \) is the histogram of oriented gradients computed in the image region, \( B \), and \( b \) and \( s \) are the orientation bin and sub-region number. The DHOG is computed by

\[
\text{DHOGB}_{s, b} = \text{HOGB}_{s, b} - \text{HOGB}_{s, b - 1}, \tag{8}
\]

where \( k \) is a scale parameter. Therefore, each scaled HOG can be computed by

\[
\text{HOGB}_{s, b} = \sum_{k = s_1}^{s_2} \text{DHOGB}_{s, b}, \tag{9}
\]

where \( s_1 \) and \( s_2 \) are scale range \([s_1, s_2]\) and

![Fig. 2](image-url) Structure of delta-HOG according to each scale change.
Fig. 3  RSIRF training and testing processes.

$HOG_{B,s_1}(s, b)$ is equal to $DHOGB_{B,s_1}(s, b)$. In Fig. 2, $s_1$ is 8, and $s_2$ is 24. The DHOG is very useful because we can calculate only the case with the largest patch window, the other cases with smaller patch windows are not calculated, but can be reused from those of the largest patch window, which decreases a lot of the computational cost.

2.2 Cloth Pattern Recognition Process Using Rotation and Scale Invariant Random Ferns

As shown in Fig. 3, rotation and scale invariant random ferns (RSIRF) using both the RIHLBF descriptor and the DHOG is proposed. Only grey level pixel data are used for comparing the proposed method with dense-SIFT. The FAST-9 algorithm is used for detecting the keypoints from all training cloth patches and when the patches have only edges without corners, the FAST-9 threshold parameter is adjusted to detect an edge instead of a corner. The RIHLBF descriptors are generated in each keypoint according to the scale level. After acquiring them, 250 keypoints (i.e., words) are created by K-means clustering. After this, all of the class specific training data are trained by random ferns to generate model data which have probability distributions of classes.

The testing (query) step is similar to the training step, FAST-9 is used and the RIHLBF descriptors are extracted from the query (test) images, and classified. From the database in the bag of words (BoW) [11], it then selects a similar patch image that has a high probability of being similar to the query patch image.

3. Experimental Results

In this experiment, we used a database with a total of 800 images, formed at a constant-size of 120 x 120 pixels, from which the training is carried out. Among the images in the database, the 10 cloth patch data from each of the 36 clothing types, specially captured with various pose and scale changes, are included. The proposed method is compared with dense-SIFT. In the dense-SIFT process, a 32 x 32 pixel window moves on the training image in a 4-pixel interval, and by using these SIFT vectors, K-means clustering and the binary K-D tree are used.

As shown in Fig. 4, the proposed method won 20 out of 36 test cases of which both the proposed method and dense-SIFT were used to find the 5 most similar cloth patches from the database. In addition, the proposed method won 18 out of 36 test cases in which both were used to find the 10 most similar cloth patches.

Figure 5 shows the searching results for the case in which the proposed method is better than the dense-SIFT method. The top-left patch image is the most similar result image; next to it on the right is the second to the most similar one, and the bottom-right patch is the least similar one. The query images for Q16, Q22, and Q27 tests are the top-left patch images in Fig. 5 (a), (c), and (e). Plaid and dot pattern-based cloth patches are well detected because the proposed DHOG-based RIHLBF descriptors acquired from small to large scaled correspondences are all trained to discriminate a larger number of various poses and scale changes than dense-SIFT. As can be seen in Fig. 5 (a), (c) and (e), each patch that has rotation and scale changes is well detected by the proposed method. Even though these patches have some wrinkles, the proposed method exhibits better performance than dense-SIFT. As shown in Fig. 5 (d), dense-SIFT does not detect the same patch image as the query image, and as shown in Fig. 5 (b), (d), and (f), only a few similar patch images are detected.

Figure 6 presents search results for a case in which
Fig. 6 Comparing result image between the proposed and Dense-SIFT in the case that Dense-SIFT is better.

Table 1 Time consumption comparison between the proposed method, RSIRF, and Dense-SIFT.

| Case   | Method         | Time (sec) | Difference |
|--------|----------------|------------|------------|
| Training | Proposed       | 0.11/image | 57.7%      |
|         | Dense-SIFT     | 0.26/image |            |
| Testing | Proposed       | 0.17       | 41.4%      |
|         | Dense-SIFT     | 0.29       |            |

the dense-SIFT method is better than the proposed method. Similarly, the top-left patch image is the most similar one and bottom-right patch is the least similar one. The proposed method is worse in the case of flower pattern-based cloth patches. As can be seen in Fig. 6 (a) and (c), when these query patches have a complex pattern, the RSIRF has difficulty with finding similar patches that have pose and scale changes. This is because the proposed method loses most of the pixel location information; therefore it has some limitations with discriminating detailed patterns such as this. Increasing the number of sub-regions can make it better, but pose and noise invariant characteristics will be worse. These have a trade-off relationship.

The training and testing processing time comparison results are shown in Table 1. When compared with dense-SIFT on a PC with an Intel Xeon 2 GHz CPU and 16 GB RAM, the proposed method decreases the running time by 57.7% in training, and 41.4% in testing.

4. Conclusions

In this paper, a rotation and scale invariant DHOG based random ferns method is proposed for cloth pattern recognition with high performance. While the existing SIFT method has difficulty with cloth pattern recognition due to the deformable nature of the training cloth image which has a lot of pose and scale changes, the proposed method overcame these disadvantages and won 20 test cases out of a total of 36 test cases in which the proposed method and dense-SIFT were used to find the 5 most similar cloth patches, respectively. In addition, the proposed method cut the computation time down by 57.7% in training, and 41.4% in testing, which is quite helpful for improving cloth pattern recognition which requires fast processes for both training and testing.

Acknowledgments

This research was supported by the Ministry of Science, ICT and Future Planning (MSIP), Korea, under the Information Technology Research Center (ITRC) support program (NIPA-2014-H0301-14-1012) supervised by the National IT Industry Promotion Agency (NIPA)

References

[1] X. Yang, S. Yuan, and Y.L. Tian, “Recognizing clothes patterns for blind people by confidence margin based feature combination,” Proc. 19th ACM international conference on Multimedia, pp.1097–1100, 2011.
[2] M. Yang and K. Yu, “Real-time clothing recognition in surveillance videos,” Image 2011 18th IEEE International Conference on Processing (ICIP), IEEE, pp.2937–2940, 2011.
[3] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, “Local features and kernels for classification of texture and object categories: A comprehensive study,” International journal of computer vision, vol.72, no.2, pp.213–238, 2007.
[4] M. Pietikäinen, T. Ojala, and Z. Xu, “Rotation-invariant texture classification using feature distributions,” Pattern Recognit., vol.33, no.1, pp.43–52, 2000.
[5] S. Yuan, Y.L. Tian, and A. Arditi, “Clothing matching for visually impaired persons,” Technology and Disability, vol.23, no.1, pp.75–85, 2011.
[6] D.G. Lowe, (1999). “Object recognition from local scale-invariant features,” Proc. International Conference on Computer Vision 2, pp.1150–1157.
[7] M. Ozsysal, P. Fua, and V. Lepetit, “Fast keypoint recognition in ten lines of code,” IEEE Conference on Computer Vision and Pattern Recognition, pp.1–8, 2007.
[8] M. Villamizar, F. Moreno-Noguer, J. Andrade-Cetto, and A. Sanfeliu, “Efficient rotation invariant object detection using boosted random ferns,” 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.1038–1045, 2010.
[9] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection.” IEEE Conference on Computer Vision and Pattern Recognition, pp.886–893, 2005.
[10] C. Bibby and I. Reid, “Fast feature detection with a graphics processing unit implementation,” Proc. International Workshop on Mobile Vision, 2006.
[11] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, “Visual categorization with bags of keypoints.” Workshop on statistical learning in computer vision, ECCV, vol.1. no.1-22, pp.1–2, 2004.