CATEGORY-SEPARATING STRATEGY FOR BRANDED HANDBAG RECOGNITION

Yan Wang, Sheng Li, Alex C. Kot

School of Electrical and Electronic Engineering
Nanyang Technological University
Singapore 639798

ABSTRACT

In recent years, computer vision community has devoted efforts on the recognition of basic-level categories. On the other hand, fine-grained object recognition which targets at recognizing objects belonging to the same basic-level class, is a more challenging problem and receives an increasing attention during recent years. In this paper, we propose a hierarchical structure Category-Separating Strategy for branded handbag recognition which is the first attempt to address the fine-grained object recognition on branded handbags. Experimental results on a newly constructed dataset are provided to show the effectiveness of the proposed methodology.

Index Terms— fine-grained, object recognition, handbag dataset

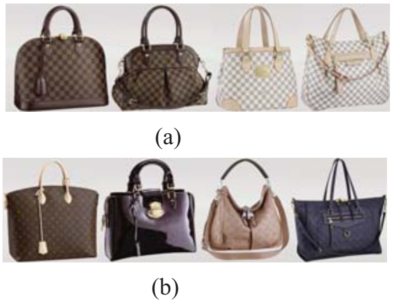
1. INTRODUCTION

Unlike basic-level recognition (*e.g.*, differentiate between car and house), fine-grained categorization problem targets on distinguishing subordinate-level categories with subtle differences. Due to its challenge and practicability, many researchers are devoted to investigating it over the past several years. There are some well-studied cases of fine-grained visual categorization in multiple application domains such as the species or breed recognition [1][2][3][4][5][6], human activity recognition [7], food recognition [8][9], as well as some fashion items analysis [10].

Domain specific knowledge is explored in many fine-grained categorization approaches. For example, bird species are identified by their unique shapes, appearances or poses. Researches concerning about fashion like clothes always analyze clothes attributes, human pose and spatial positions of fashion items [10].

Our target is different from those existing fine-grained categorization techniques. In this paper, we try to recognize a handbag that belongs to a specific brand. The motivation is the need of the manufactures to gain a deeper insight into the customer's feedback of their handbags, which are usually uploaded (as images) on blogs without the specific model information. To the best of our knowledge, there is no prior published work on the branded handbag.

To address this problem, we propose a Category-Separating Strategy, which describes the attributes of a handbag and provides a hierarchical framework for recognition. We identify a specific brand of handbags into two basic categories, i.e., handbags with checkerboard pattern and without checkerboard pattern. Handbags can be further recognized within each category. We construct a dataset consisting of 80 branded handbags from the website of a world-renowned brand [11]. Experimental results show that proposed Category-Separating Strategy can boost the recognition accuracy.


(a)

(b)

Fig. 1. Examples of two categories of the specific brand. (a) checkerboard pattern; (b) non-checkerboard pattern.

2. HANDBAG CATEGORY-SEPARATING STRATEGY

We observe that handbags with or without checkerboard pattern are the two basic categories for the specific brand (i.e., Louis Vuitton Handbag), as shown in Fig.1. Checkerboard pattern is a distinguished attribute for the branded handbag recognition.

We design a methodology to differentiate checkerboard pattern vs. non-checkerboard pattern at the first stage of hierarchical structure. Images of checkerboard pattern in a handbag always suffers from distortions such as pleated structure, illumination changes or rotation. General methods cannot work well in such circumstances. We are inspired by the work of Geiger *et al.* [12], which is the state-of-the-art checkerboard corner detector. Their method is proven to be useful for detecting checkerboard corners even with heavy distortion.

At the beginning, Geiger *et al.* [12] compute a corner likelihood of each pixel. A list of corner candidates is derived based on the likelihood and two dominant edge orientations of the corner candidates are estimated afterwards. The work in [12] designs a checkerboard recovery procedure to identify the existence of the checkerboard pattern. According to our observation, the recovery procedure would fail when the checkerboard pattern displays irregularly. We propose a product-based local checkerboard re-validation method for these corners, where two newly designed weight matrices are proposed to verify whether a corner candidate is a checkerboard corner.

This research is supported by the Singapore National Research Foundation under its IDM Futures Funding Initiative and administered by the Interactive & Digital Media Programme Office, Media Development Authority.

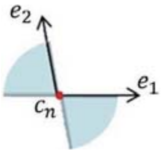


Fig. 2. One of the ideal checkerboard patterns of the candidate corner $\mathbf{c}_n$, which computed from the two dominant orientations e_1 and e_2.

2.1. Weight matrix design

Let $\mathbf{C} = \{\mathbf{c}_1, ..., \mathbf{c}_n, ..., \mathbf{c}_N\}$ denote the set of corner candidates of a handbag image I (computed using the method in [12]), where each entry in $\mathbf{C}$ indicates the sub-pixel location of nth corner candidate and $\mathbf{c}_n = (x_n, y_n)$, N refers to the number of the corner candidates. We define a local region L_n for a corner candidate $\mathbf{c}_n$ as a circle centered at (x_n, y_n) with the radius

$$r_n = \kappa(\min_{\mathbf{c}_m \in \mathbf{C}, \mathbf{c}_m \neq \mathbf{c}_n} ||\mathbf{c}_m - \mathbf{c}_n||_2) \tag{1}$$

where $|| \cdot ||_2$ is the l_2 norm and κ is a constant, determining how much portion of the distance between corner candidate $\mathbf{c}_n$ and its nearest corner candidate should be taken as the radius. The value of κ should be within the range of $[0, 1]$. We experimentally set $\kappa = \frac{2}{3}$ to get sufficient neighborhood information (if κ is approximate to 0, then most neighboring information would be lost) and alleviate the errors of some undetected neighboring corner candidates (if it is approaching 1, it is lack of the ability to handle distortion or missed candidates).

For a candidate corner $\mathbf{c}_n$, we adopt the normal probability density function to compute a local weight matrix $\mathcal{W}_n$ as follows

$$\mathcal{W}_n(\mathbf{q}) = f\{||\mathbf{q} - \mathbf{c}_n||_2 | \mu_n, \sigma\} \tag{2}$$

where $\mathbf{q} = (x, y)$ is the location of a pixel inside L_n, $\mu_n = (x_n, y_n)$ and $\sigma = \frac{r_n}{2}$.

2.2. Checkerboard validation

We check whether a candidate corner c_n is a checkerboard corner by comparing the pixel pattern in the local region L_n with the ideal checkerboard patterns. Such ideal patterns can be computed based on the two dominant orientations estimated from $\mathbf{c}_n$ (please refer to [12] for the dominant orientation estimation). Fig. 2 gives an example of an ideal checkerboard pattern (a binary image) computed for a candidate corner $\mathbf{c}_n$, where the two estimated dominant directions are e_1 and e_2, respectively. For simplicity, we define such a pattern as $\mathcal{M}_{n1}$ for corner c_n. There are two ideal checkerboard patterns for one candidate corner, i.e., $\mathcal{M}_{n1}$ and $\mathcal{M}_{n2} = \overline{\mathcal{M}_n}$, where $\overline{()}$ refers to the *not* operator.

To facilitate the comparison between the pixel pattern in L_n and the ideal checkerboard patterns which are binary, we binarize the pixels in L_n as follows.

$$I'(\mathbf{q}) = \left\lfloor \frac{I(\mathbf{q}) - \min\limits_{\mathbf{q} \in L_n}\{I(\mathbf{q})\}}{\max\limits_{\mathbf{q} \in L_n}\{I(\mathbf{q})\} - \min\limits_{\mathbf{q} \in L_n}\{I(\mathbf{q})\}} + 0.5 \right\rfloor \tag{3}$$

where $\mathbf{q} \in L_n$ refers to a pixel inside L_n, $I(\mathbf{q})$ is the intensity of the pixel $\mathbf{q}$ and $\lfloor x \rfloor$ denotes the largest integer that does not exceed x. The difference between the pixel pattern in L_n and the two ideal checkerboard patterns are then measured as

$$D_i = \frac{\left| \sum\limits_{\mathbf{q}} I'(\mathbf{q}) \cdot \mathcal{W}_n(\mathbf{q}) - R_{ni} \right|}{R_{ni}} \tag{4}$$

where $\mathbf{q} \in L_n$, $i = 1, 2$, and $R_{ni} = \sum\limits_{\mathbf{q}} \mathcal{M}_{ni} \cdot \mathcal{W}_n$ which is computed from the two ideal checkerboard patterns. A candidate corner $\mathbf{c}_n$ will be verified as a checkerboard corner if the following two conditions are satisfied: i) max$(D_1, D_2) > \xi$ and ii) min$(D_1, D_2) < 1 - \xi$, where $\xi \in [0, 1]$. The closer the ξ is to 1, the more strict the rule is to verify the corner candidate as a checkerboard corner. A slightly smaller ξ offers a room for noise and distortion, which is set as 0.8 in our implementation. A handbag image can be identified as the handbag with checkerboard pattern if the number of verified checkerboard corners beyonds a preset threshold T.

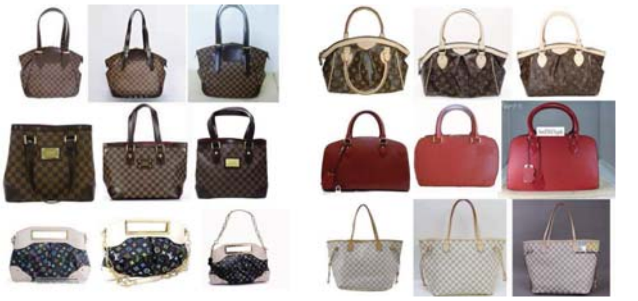


Fig. 3. Examples of handbag dataset.

3. EXPERIMENTAL RESULTS

3.1. Handbag dataset construction

According to our investigation, there are 551 different handbags for women on official website of this branded handbag [11]. Our dataset is a collection over 976 images covering 80 representative handbags with different style names. Number of images for each handbag varies according to the amount of the sources from the internet. Fig. 3 shows some examples of our dataset. Noted that like the attribute discovery work of [13], the images in our dataset have relatively clean background, allowing us to focus on the handbag itself. Thus the recognition will not be influenced by some visual challenges like clutter background or occlusion. In addition, the images in our dataset contain mainly the frontal view of handbags. We believe that for a handbag, if users desire to recognize it, they always pay more attention to its frontal view rather than side views or even the reversed side. We will build a large dataset concerning all handbags from this brand in the future. More complicated situations will be considered, such as clutter background, with some degree of 3-d view-point rotation.

3.2. Benchmarks

In the first part of our experiment, we evaluate the accuracy of some standard computer vision recognition algorithms on the handbag dataset [14]. We extract three popular image features: SIFT feature [15] in dense grid, LBP texture features [16] and color histograms

Table 1. Recognition accuracy of single features and concatenation of those features.

image features	*classification accuracy*
SIFT + Bag-of-Words	68.62%
LBP + Bag-of-Words	32.74%
SIFT&LBP + Bag-of-Words	58.39%
SIFT&Color + Bag-of-Words	68.21%
LBP&Color + Bag-of-Words	33.56%
SIFT&LBP&Color + Bag-of-Words	58.53%

Table 2. Comparison of different identification results.

methods	*identification accuracy*
SIFT + Bag-of-Words	96.31%
LBP + Bag-of-Words	90.41%
chekerboard detection[12]	97.10%
proposed	**99.59%**

[17]. SIFT feature is extracted from densely located patches centered at every 5 pixels in the image with a patch size of 7×7 pixels. A 58-dimentional LBP feature is extracted densely from the patch size of 10×10 pixels. A 1000-word dictionary together with localized soft assignment [18] encoding and three level spatial pyramid with max-pooling are used for both the two types of features. 11-dimentional color histograms are extracted by using color naming methods proposed by [17].

We employ the following strategy for all the experiments. Three images per handbag are used for training and the rest are used for testing. We train different multi-class SVM classifiers [19] using different types of features as well as their possible combinations.

Table 1 summarizes the recognition results by using different features. As can be seen from the result that SIFT alone could obtain a higher accuracy than LBP, which means that SIFT is more suitable than LBP in the representation of bag texture. Because SIFT feature receives a relatively higher results, it would be treated as the baseline for handbag recognition.

3.3. Checkerboard pattern identification

To demonstrate the value and applicability of proposed Category-Separating Strategy, we compare different methods for handbag checkerboard pattern identification. As indicated in section 2.2, a handbag will be identified as checkerboard pattern if the number of the verified checkerboard corners is over different threshold values (T). Fig. 4 shows the identification accuracy under different T over all the handbags in dataset. It can be seen that $T = 7, 8, 9$ or 10 gives the best performance with an accuracy of 99.59%. We also extract SIFT and LBP features from a handbag image for the checkerboard pattern identification, respectively. The same aforementioned bags of words, encoding and pooling methods are used. In the identification experiment, we use three images per handbag for training and rest images for testing. Different binary-class SVM classifiers are trained for SIFT and LBP respectively. Table 2 gives the comparison results. It is important to note that our method achieves the highest performance without training procedure.

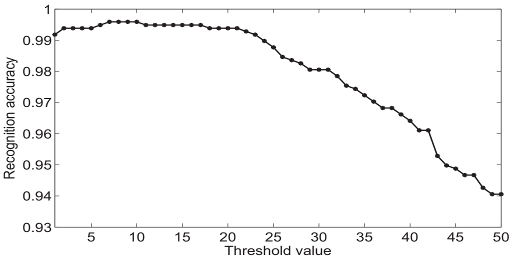


Fig. 4. Performance of the checkerboard pattern identification.

Table 3. The performance of the two types of evaluations for the handbag recognition.

methods	*recognition accuracy*
Type A	68.62%
Type B	**71.76%**

3.4. Handbag recognition

As mentioned before, we adopt the SIFT feature in dense grid with the bags of words, followed by multi-class SVM as baseline recognition technique for the handbags. We conduct the following two types of evaluations.

- Type A: we train a classifier using the baseline method for all the handbags in our dataset. In the testing, we compare a handbag against all the handbags in the dataset using the classifier.
- Type B: we train a classifier using the baseline method for each of the two categories. In the testing, we first identify which category a handbag belongs to using our proposed Category-Separating Strategy. Then this handbag is compared against all the handbags within the identified category using the classifier.

Table 3 gives the comparison results between these two types of recognition. It can be seen that the recognition accuracy yields a 3.14% gain for the overall system by exploiting Category-Separating Strategy.

4. CONCLUSIONS

This paper is the first to address branded handbag recognition which is a practical but challenging problem. We propose a Category-Separating Strategy to enable recognition on separate categories. We construct a handbag dataset containing 80 models of representative branded handbags. The experiments on this dataset shows the effectiveness of the proposed Category-Separating Strategy for the branded handbag recognition. In the future, we will expand the dataset to a sufficiently large size to include all the handbags of the same brand.

5. REFERENCES

[1] J. Deng, J. Krause, and L. Fei-Fei, “Fine-grained crowdsourcing for fine-grained recognition,” in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, 2013, pp. 580–587.

[2] O. Parkhi, A. Vedaldi, A. Zisserman, and C. Jawahar, “Cats and dogs,” in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 3498–3505.

[3] R. Farrell, O. Oza, N. Zhang, V. Morariu, T. Darrell, and L. Davis, “Birdlets: Subordinate categorization using volumetric primitives and pose-normalized appearance,” in *Computer Vision (ICCV), 2011 IEEE International Conference on*, 2011, pp. 161–168.

[4] S. Branson, C. Wah, F. Schroff, B. Babenko, P. Welinder, P. Perona, and S. Belongie, “Visual recognition with humans in the loop,” in *Proceedings of the 11th European conference on Computer vision (ECCV)*, 2010, pp. 438–451.

[5] G. Martinez-Munoz, N. Larios, E. Mortensen, W. Zhang, A. Yamamuro, R. Paasch, N. Payet, D. Lytle, L. Shapiro, S. Todorovic, A. Moldenke, and T. Dietterich, “Dictionary-free categorization of very similar objects via stacked evidence trees,” in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 549–556.

[6] M.-E. Nilsback and A. Zisserman, “A visual vocabulary for flower classification,” in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 2, 2006, pp. 1447–1454.

[7] B. Yao, A. Khosla, and L. Fei-Fei, “Combining randomization and discrimination for fine-grained image categorization,” in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 2011, pp. 1577–1584.

[8] F. Kong and J. Tan, “Dietcam: Regular shape food recognition with a camera phone,” in *Body Sensor Networks (BSN), 2011 International Conference on*, 2011, pp. 127–132.

[9] S. Yang, M. Chen, D. Pomerleau, and R. Sukthankar, “Food recognition using statistics of pairwise local features,” in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, 2010, pp. 2249–2256.

[10] K. Yamaguchi, M. Kiapour, L. Ortiz, and T. Berg, “Parsing clothing in fashion photographs,” in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 3570–3577.

[11] L. Vuitton, http://www.louisvuitton.eu/front/\#/engE1/Collections/Women/Handbags.

[12] A. Geiger, F. Moosmann, O. Car, and B. Schuster, “Automatic camera and range sensor calibration using a single shot,” in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, 2012, pp. 3936–3943.

[13] T. L. Berg, A. C. Berg, and J. Shih, “Automatic attribute discovery and characterization from noisy web data,” in *Proceedings of the 11th European conference on Computer vision: Part I*, 2010, pp. 663–676.

[14] M. Chen, K. Dhingra, W. Wu, L. Yang, R. Sukthankar, and L. Yang, “Pfid: Pittsburgh fast-food image dataset,” in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, 2009, pp. 289–292.

[15] D. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[16] T. Ojala, M. Pietikainen, and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, 2002.

[17] J. van de Weijer, C. Schmid, J. Verbeek, and D. Larlus, “Learning color names for real-world applications,” *Image Processing, IEEE Transactions on*, vol. 18, no. 7, pp. 1512–1523, 2009.

[18] L. Liu, L. Wang, and X. Liu, “In defense of soft-assignment coding,” in *Computer Vision (ICCV), 2011 IEEE International Conference on*, 2011, pp. 2486–2493.

[19] C.-C. Chang and C.-J. Lin, “LIBSVM: A library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.