Product Recognition: Survey

Turkish Online Journal of Qualitative Inquiry (TOJQI)

Volume 12, Issue 9, August 2021: 5496 – 5516

Product Recognition: Survey

Yousef Alsahafi

University of Jeddah

ysalsahafi@uj.edu.sa

Abstract

Product recognition systems have attracted wide attention because they help guests to gain the product applicable information veritably presto [e.g., product's price, the nearest store dealing this product, consumer recommendations products,etc.] With the recent growth of the space, this is the first check of product recognition systems. In addition to a check of the many living workshops, we bandy openings and open problems.

The idea of this composition is to give new experimenters a brief and scrutable review of product recognition approaches. We bandy product recognition on mobile device systems that use the camera-phones. Also, cloth-ing recognition styles calculate on the estimated mortal disguise. Grocery products are honored using original features or an image bracket approach. We also present the datasets that have been used in product recognition operations. Eventually, we give experimenters open problems and advice.

In this paper, we prefer to give a general picture of object discovery approaches. Since A product is an object, we assume the approach is used to expostulate discovery and recognition may be used to product discovery and recognition. First, what's object discovery? Then, we have a fixed number of object orders and try to find all cases of *these orders inside the image and also draw bounding boxes on them. Object discovery was counting on SIFT and HOG descriptors for a period of time. They're low-position features. To achieve the stylish performance on object discovery experimenter combined multiple low-position im- age features with a high-position environment. Still, Object discovery performance has metamorphosed during 2010-2016.*

Keywords: *consumer, pattern recognition, R-CNN,*

Introduction

We proposed that now after we've seen the most notorious models of object discovery, let’s move to our content that's product recognition and show how that's related to object discovery and recognition. Product recognition is a process of assaying image queries of marketable products to find out applicable in- conformation [e.g., price, near a store, consumer recommendations,etc.]. It's becoming an active area in computer vision. While there are myriad implicit operations, we've planted three types of product recognition systems. The first one is product recognition on a mobile device, which uses a camera-phone to take a picture of a product and also returns some information of the product[19,11].

Reclamation systems but take time between taking a capture and the response to the query. They take knockouts of seconds. Product recognition on mobile device systems has faced some challenges. For case, they should reduce the quantum of data that will be transferred from the mobile to the garcon to reduce the calculation time.

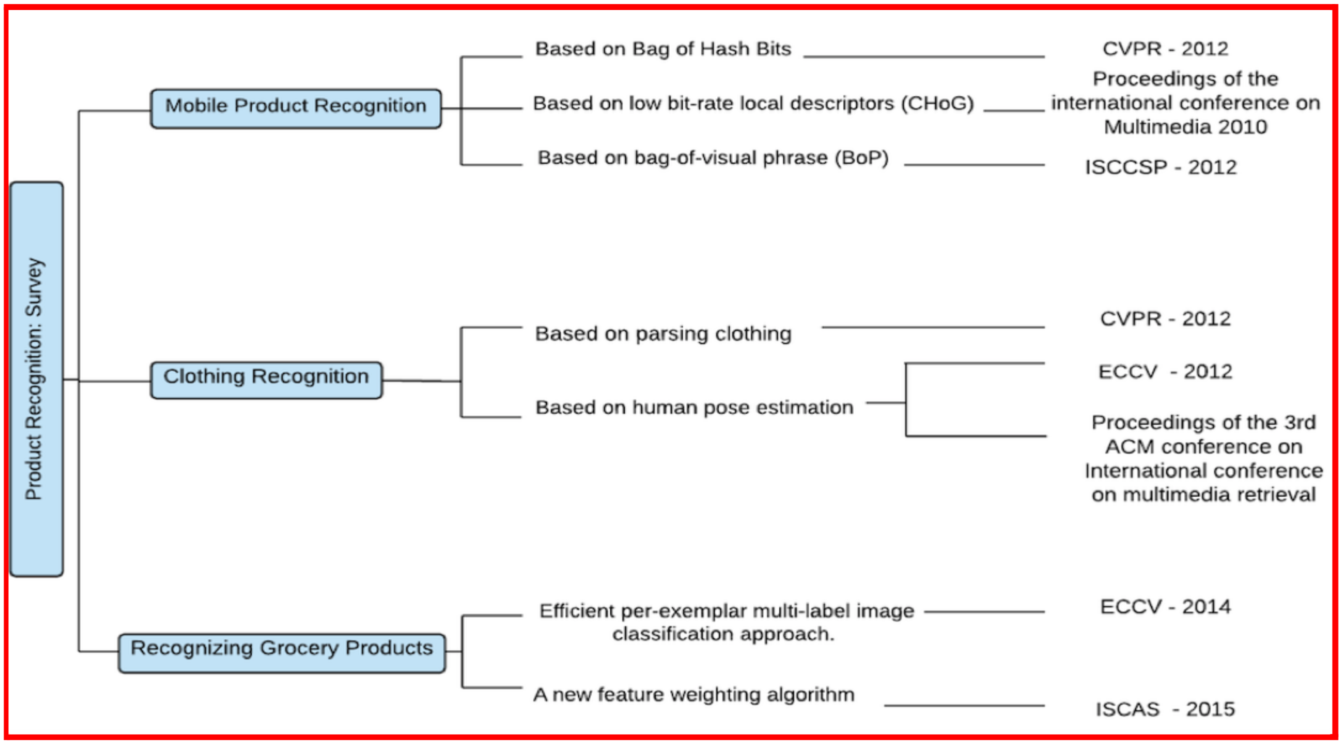

Product Recognition: Survey
Mobile Product Recognition
Based on Bag of Hash Bits
CVPR - 2012
Based on low bit-rate local descriptors (CHoG)
Proceedings of the international conference on Multimedia 2010
Based on bag-of-visual phrase (BoP)
ISCCSP - 2012
Clothing Recognition
Based on parsing clothing
CVPR - 2012
Based on human pose estimation
ECCV - 2012
Proceedings of the 3rd ACM conference on International conference on multimedia retrieval
Recognizing Grocery Products
Efficient per-exemplar multi-label image classification approach.
ECCV - 2014
A new feature weighting algorithm
ISCAS - 2015

Figure 1. Overview of product recognition survey

The alternate type of product recognition system is cloth-ing recognition, which recognizes clothes on a person. Utmost apparel recognition systems use mortal acts estimation ways to descry the person and also fete the clothes worn by this person [3]. Clothing recognition approaches also have faced some challenges. For in-station, if the disguise estimation isn't correct, the results of the apparel recognition system will frequently not be correct. Also, clothes with periodic color changes [e.g, stripes] can beget an incorrect disguise estimation which reduces the accuracy of the apparel recognition fashion.

The third type of product recognition is feting grocery products. There are numerous products that have analogous package shapes or colors, which makes recognition harder. Another thing that makes fitting grocery products delicate is that the test images are taken in fully different settings than the training images and these images suffer from blur and veritably different lighting conditions.

To the stylish of our knowledge, there's no specific check for product recognition. To fill this gap, we give a comprehensive overview of product recognition in this paper, including the following product recognition on a mobile device, apparel recognition, and grocery products recognition. The donation of our work is twofold [1] Wepro- vide a comprehensive check of product recognition. [2] We give open problems and some analysis advice to re-quest.

Figure 1 shows the three types of product recognition systems. Also, it shows each type uses further than one fashion to fete the product. For case, product recognition on a mobile device has used three different approaches. The three approaches aim to ameliorate the performance of the mobile visual hunt. He et al. [17] have achieved high performance of mobile visual hunt grounded on Bag of Hash Bits fashion. The alternate type is apparel recognition, which also uses different styles to fete garments on a person. Kalantidis etal. [19] show an approach that's 50 times faster than state-of-the-art apparel detection. Their approach relies on the mortal disguise estimation. The last type is grocery product recognition, which has two approaches to fete grocery products. One of them can fete one product from the query image and the alternate bone can fete further than one product from the query image.

Object Discovery

During the last decade, SIFT [2] and Overeater [6] were notorious descriptors and have been used a lot on visual recognition tasks. SIFT descriptors are considered one of the stylish algorithms in computer vision to descry and describe original features in images. It was published by David Lowe in 1999 [2]. It could induce a large number of features and also is steady to gyration and metamorphosis. The SIFT system transforms an image into numerous original point vectors; each vector is steady to image restatement, scaling, and gyration. Histogram of Acquainted Grade [Overeater] descriptors are point descriptors used in Computer Vision for the discovery of the object at the image. Overeater descriptions bluffs have been introduced by Navneed Dalal and Bill Triggs in 2005 [6].

Overeater descriptors algorithm divides the image into small connected regions that are called cells and also compiling a histogram of grade direction for each cell. The descriptor is the combination of these histograms.

They helped to make progress on colorful visual recognition tasks. Experimenters realized the progress of the object discovery performance has come slow during 2010-2012. Krizhevsky etal. [18] have shown how a CNN makes a huge difference on image bracket delicacy on the ImageNet Large Scale Visual Recognition Challenge [ILSVRC] [7, 8]. Their results make Girshick etal. [15] supposed to use discovery as a bracket. R-CNN proposition is to make ground to fill the gap between image classification and object discovery. By making this connection their approach could achieve a mean average perfect which is better than the former stylish result by 30 on PASCAL VOC 2012.

For illustration, CNN generates 4096-dimensional features while in the UVA discovery system, they produce 360k-dimensional features. R-CNN needs 13s/ image on a GPU or 53s/ image on a CPU to compute region proffers and features. The only calculations they need are when they need to be class-specific and do fleck products between features and SVM weights and non-maximum suppression. To give further detail about their matrices, the point matrix is generally 2000 x 4096 and the SVM weight matrix is 4096 x N, where N is the number of classes. With ultra-modern multi-core CPU matrix proliferation of these mamas- beats will be presto. Since, they've low-dimensional features, need1.5 GB, while, for illustration, UVA system requires 134 GB due to its high-dimensional features.

Indeed, the R-CNN has downsides. First, it's veritably slow at test- time, since it has 2k image regions and runs each region on CNN. Also we want to estimate CNN on each re-gion that makes R-CNN veritably laggardly at test time. R-CNN uses a Picky Hunt system to prize region proffers while EdgeBoxes generates the stylish offer quality and faster than Picky Hunt system (0.2 seconds per image). Also, the R-CNN model has the multistage training channel that's complex.

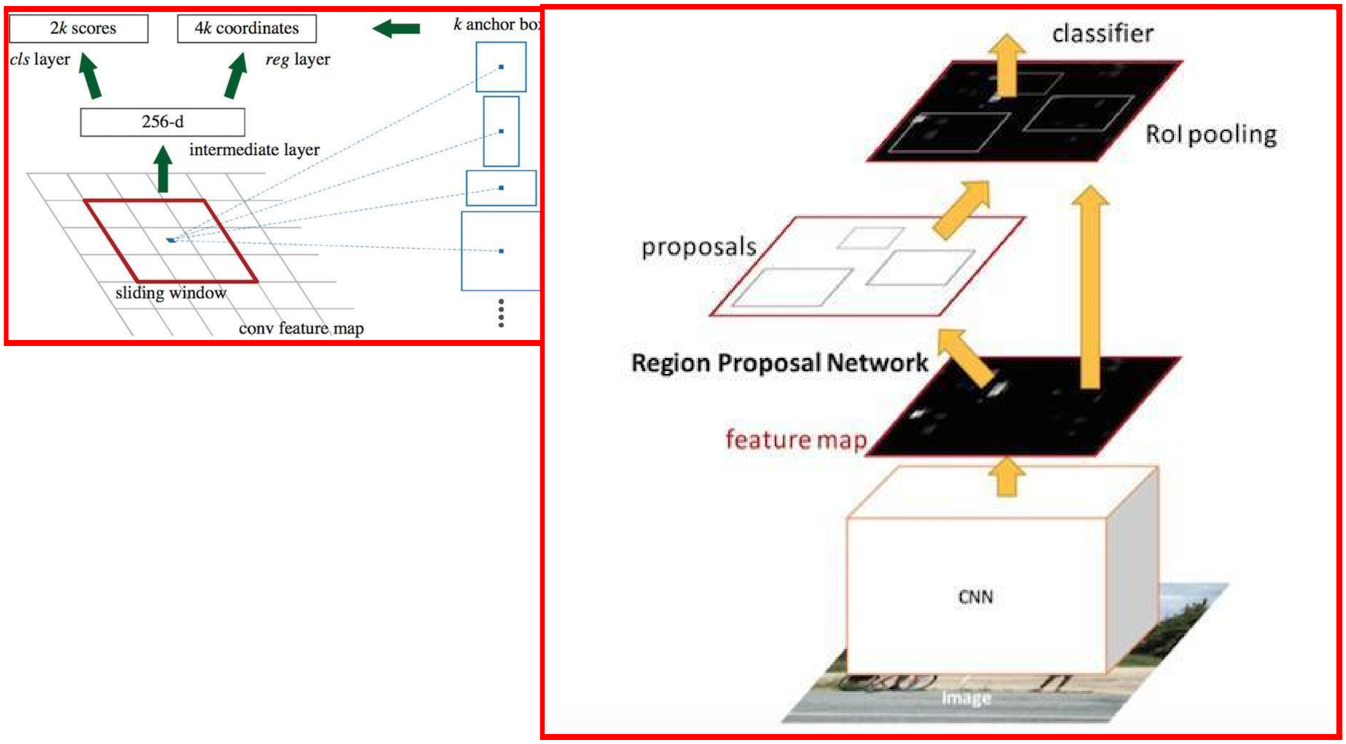


Figure 2: Proposal classification

The convolutional layers are the RPN and contemporaneously regress region bounds and objectness scores at each position on a regular grid. So, the RPN contains many fresh convolutional layers and that shows the RPN is a kind of completely convolutional network (FCN) and can be trained end-to-end.

They propose a training model to combine RPNs with Fast R-CNN [2] object discovery network. The training model unified fine-tuning for RPNs and also fine-tuning for object discovery (Fast R-CNN). This model allows making a network with convolutional features that are participated between both tasks. Figure [1] shows their entire system.

Using discovery as retrogression means, your input is an image and your affair is figures. These figures are coordinated for bounding boxes that are drowned around the object. Their paper is called You Only Look Formerly (YOLO) Unified, Real-Time Object Discovery). YOLO is a single neural network that predicts bounding boxes and class probabilities from the full images.

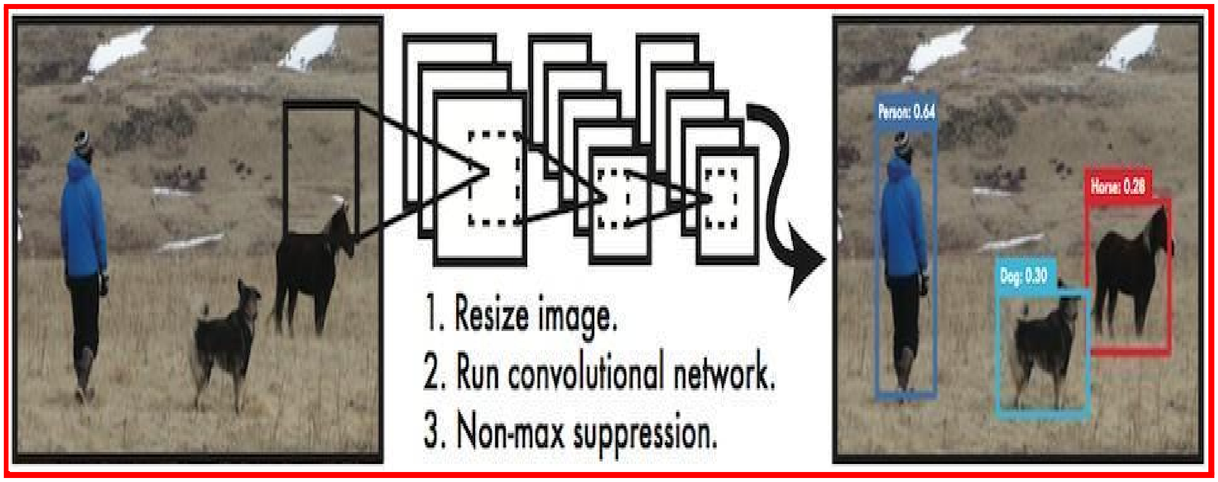


Figure 3.: Classification by the object detection

here is classification with different object detection are shown in figure 3
The spatial constraint on bounding box predictions limits the number of nearby objects that YOLO can predict. Since YOLO allows each grid cell to only predicts two boxes and can only have one class. Their model has difficulty with small objects that appear in groups, such as flocks of birds.

Since our model learns to predict bounding boxes from data, it struggles to generalize to objects in new or unusual aspect ratios or configurations. Our model also uses relatively coarse features for predicting bounding boxes since our architecture has multiple downsampling layers from the input image. Finally, while we train on a loss function that approximates detection performance, our loss function treats errors the same in small bounding boxes versus large bounding boxes. A small error in a large box is generally benign but a small error in a small box has a much greater effect on IOU. Our main source of error is incorrect localizations.

Product Recognition on a Mobile Device

Product recognition on a mobile device is an important application because of the commercial importance and wide user demands. Many different mobile systems have been developed many times by researchers and there are many commercial systems on mobile product search such as Goggles, Snaptell, and Point and Find. Products are often 3D objects that have different shapes, structures, or viewpoints, as shown in Figure 3. Furthermore, it is too hard to distinguish foreground products of interest from background clutter. These points make product recognition harder. In addition, mobiles have limited memory that cannot compare to conventional desktops. While they both require searching and return the results fast.

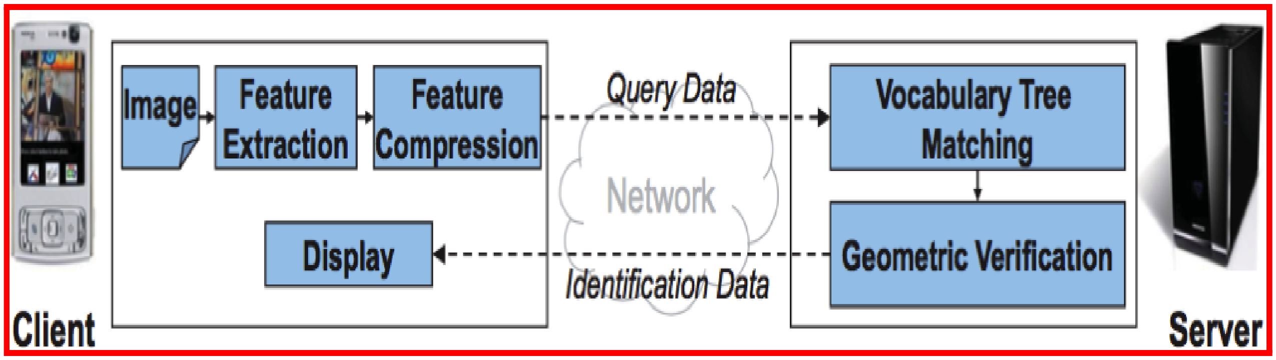


Figure 4: Product recognition on a mobile device

Figure 4. shows the architecture of the proposed product recognition on a mobile device approach based on low bit-rate local descriptors [CHoG].

approach on another dataset of different products such as food or clothing. Another limitation of this paper is that they did not compare their work with the state-of-the-art. Also, they mention that they used GV to improve the accuracy, but they did not mention which method they used. Did they use RANSAC or Hough Transform or something else?

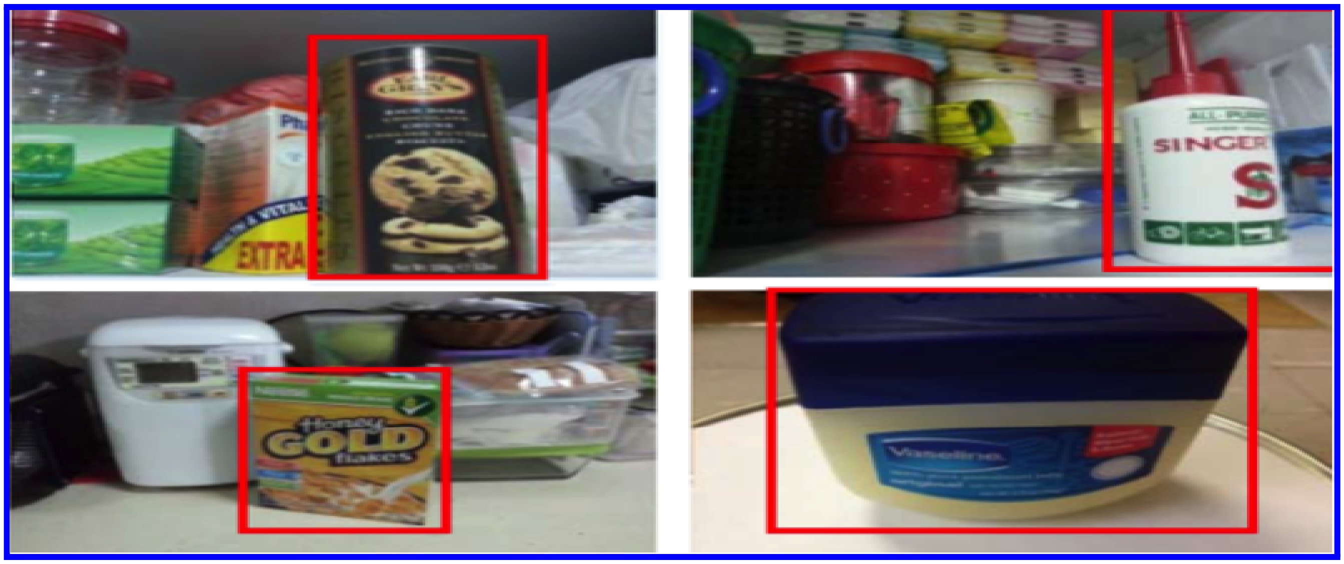


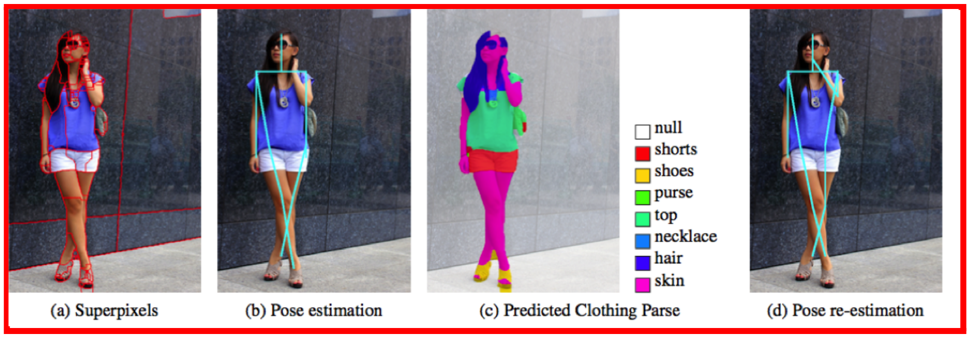
Figure 5. 3D objects that have different shapes, structures, or viewpoints.

As we mentioned early the speed is very important in product recognition on a mobile device system[14] showed another mobile application for product recognition could achieve high performance using bag-of-visual phrase [BoP]. Their system allows users to search for a product of interest by image query to find relevant information [e.g., price, nearby store, consumer recommendation, etc.]. There are many approaches that can be used to match the image query and dataset. For example, using bag-of-words [BoW] with geometric verification can improve image matching accuracy. However, geometric verification may be expensive on the online computation because the relevant image candidates will have been long-listed by the BoW method. To avoid that, researchers have used BoP that groups multiple neighboring visual words into visual phrases, and that makes visual phrases more robust against cluttered backgrounds.

Clothing Recognition

Recently, some studies have focused on clothing recognition because suggesting related clothing products to the viewer could help to increase the revenue of the market. Clothing recognition is a subset of product recognition and an extremely challenging problem because of the large number of garment items, the diversity of garment appearances, and occlusion. After reading several papers about clothing recognition, we realized most clothing recognition techniques have used human pose estimation.

However, it learns semantic attributes for clothing on the human upper body. They considered the estimation of the full-body pose from 2-D images difficult because the lower body may be occasionally occluded or otherwise not visible in some images. To this end, Yamaguchi et al. [11] have shown a robust method for parsing clothing for the whole body that im- proves on state-of-the-art models for pose estimation.



(a) Superpixels (b) Pose estimation (c) Predicted Clothing Parse (d) Pose re-estimation

Figure 6: labeling of image regions

Fashion and apparel-related issues are getting more important aspects of day to day life of people. Fashion assiduity has come a major part of the frugality, substantially over the Internet where. a large deal of business related to fashion is passing. With the rearmost specialized advancements, new fashion doors are popping up every day. Guests are also showing interest in the online shopping system so that they won't need to go to shopping promenades and spend hours buying clothes of their choice. Rather, with online shopping doors, they can simply search for their asked clothes. from anywhere using mobile or any other platform and order them. So, the online stores should. give easy ways for the guests to find their choice of clothes. In order to do that the online stores should maintain stylish quality hunt machines with all the asked features. But the problem lies with the apparel suppliers, in order to put the clothes and other particulars to these online platforms. it would bear a veritably detailed, methodical, and specific description of the particulars and apparel. suppliers aren't handed with similar means. Convolutional Neural Networks use veritably less processing compared to the rest of the image Bracket algorithms. This indicates that the network learns the pollutants which are present in traditional algorithms. The major advantages of the convolutional neural networks are the independence of the network from previous knowledge and also the trouble made in point design. Numerous important algorithms which can break the image discovery and bracket problems are being developed every time.

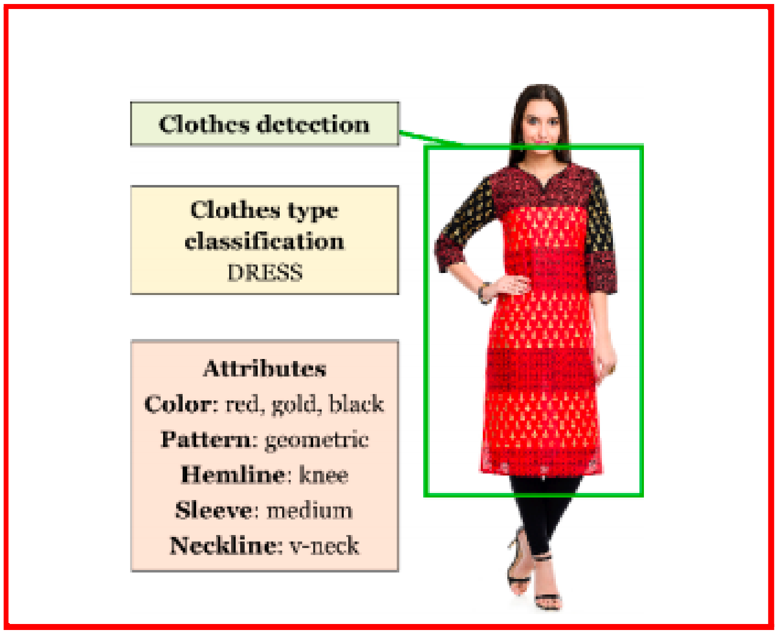


Figure 7: Clothing detection, classification & attribute assignment

As the proposed models substantially concentrate on classifying clothes in the images, it can not take occlusion into consideration. It's delicate to descry and classify clothes in images with indecorous lighting conditions and backgrounds. The position of the camera is also important for discovery of the clothes present in the images. While performing on videotape our model can not descry well if the position of person isn't facing the camera. In unborn work, the datasets can also be extended with further general types of prints taken with different types of cameras, including colorful firing angles, body positions and colorful backgrounds, both indoors and outside. There's also compass for developing new configurations for background addition which can be used to increase the data for training with colorful backgrounds. For color bracket, data for some further color clothes can be added in order to classify further types of colors.

Clothing can be used to identify the characteristics of a mortal similar as gender, age, life style etc. Since apparel plays a significant part in society, there are numerous operations related to fashion. In recent times, numerous experimenters are showing interest in colorful tasks related to apparel like feting clothes, classifying them and reacquiring them to show analogous clothes. Recent trend in new fashion related startups is the main reason for this interest in clothes- grounded recognition and discovery. These experimenters are showing interest in consumer- grounded operation i.e. apparel recognition especially related to reclamation of identical clothes grounded on patterns present on apparel particulars. Clothing analysis and understanding plays an important part in perfecting environment-apprehensive operations like automatic apparel item tagging for online shopping platforms. According to the proposition of apparel design, " Clothing stripes are determined by the combination of colorful style rudiments that parade harmonious and differentiable visual parcels". There are several problems in apparel image analysis. They're as follows

- Discovery-which includes the rough vaticination of apparel rudiments, generally indicated by blocks or places.

● Segmentation-which includes the exact localization of the rudiments (generally done pixel by pixel) in images.

● Bracket-which includes the assignment of the apparel to one of the classes present in the dataset like type of apparel.

● Trait assignment-which includes the assignment of numerous attributes like patterns, colors, styles, sizes to the apparel present in the image at formerly.

● Description-which includes the description of detected apparel present in the image in natural language.

● Retrieval-which includes searching for analogous clothes available. These problems are being dealt by experimenters since numerous times to find effective results. The results for these problems vary from hand- made complex fine models to most lately developed and largely advanced deep literacy approaches. There are 3 main types of styles used. They're as follows

● Formula grounded approach in which arbitrary apparel fine models are created manually for working apparel image analysis related problems.

● Traditional point literacy approach in which simple features like 'Histogram of Acquainted Slants' (Swillers),'Scale-Steady Point Transfigure' (SIFT)etc. are created manually and are also transferred to simple machine literacy models like' Support Vector Machines' (SVMs), Bayesian, Random Forestetc. for processing.

Datasets

Product recognition systems have been developed re- recently, and there are some datasets that have been used for product recognition. Table 1 shows the dataset for each paper in our survey.

Table 1. Product Recognition Datasets.

Author's paper	Class	Database's Description
[Tsai et al.,2010]	Product Recognition on a	They built their dataset that contains more than one million entries

	Mobile Device	which are products packaged in rigid boxes with printed labels, such as CDs, DVDs, and books.
[Zhang et al., 2014]	Product Recognition on a Mobile Device	They built their dataset, which has 41 different commercial products. The training dataset comprises 3882 reference images are obtained from cameras. While, 322 test images are taken by mobile phone with different imaging conditions such as different viewpoints, lighting, scales, cluttered backgrounds. All images are fixed to be 640 pixels for height or width.
[He et al., 2012]	Product Recognition on a Mobile Device	They built two large-scale product sets from Amazon, Zappos, and Ebay. These two datasets have various categories like shoes, clothes, groceries, electrical devices, etc. The first dataset contains 360K product images are obtained from Ama- zon.com. It comprises 15 categories of products. Also, it contains 135 test images. Each test image has one ground truth image in the dataset that has the same product as the test image. However, the product in the test image will have different object size, orientations, backgrounds, lightings, cameras, etc. The second dataset contains 400K product images that are collected for Ebay.com, Zappos.com, Amazon.com, etc. This dataset has hundreds of categories and contains 205

		test images. All images are resized to be 200-400 by 200-400 and each image con- tains about 50-500 SURF features. The authors used SaliencyCut to extract the boundaries for product objects in both dataset images and test images. These two product image sets are the largest and most challenging benchmark product datasets to date.
[Yamaguchi et al., 2012]	Clothing Recognition	They built their dataset that contains 158,235 photographs collected from Chictopia.com. Chictopia.com is website for fashion bloggers. Their dataset contains 53 diverse clothing items, of which 43 items have at least 50 image regions. There are additional labels for hair, skin, and null [background], gives a total of 56 different possible clothing la- bels. Some garment items have a large number of occurrences in the dataset such as dress [6565], bag [4431], blouse [2946], jacket [2455], skirt [2472], cardigan [1866], t-shirt [1395], boots [1348], jeans [1136], sweater [1027], etc. In addition, the dataset has a large number of items probably unheard of by the fashion non-initiate such as vest [955], cape [137], jumper [758], wedges [518], and romper [164].

[Kalantidis et al., 2013]	Clothing Recognition	This paper has used two datasets: Fashionista dataset has been used on experiments of clothing class de- tection. It conations of 685 photos. These photos are real-world photos with a human model in a cluttered but prominent position. The training dataset contains 53 different clothing labels, plus the labels hair, skin, and null. The second dataset contains of 1.2 million products from Yahoo Shopping.
[Chen et al., 2012]	Clothing Recognition	They built their dataset, which are built from Sartorialist and Flickr. They collected images with people. The dataset contains 1856 im- ages that have clothed people. Most of the images are presented on the streets.
[George et al., 2014]	Grocery Products	They have used two datasets. The first dataset contains 680 annotated test images from the proposed Grocery Products dataset, with a total number of 3235 products in 27 leaf node classes. The second dataset contains 885 test images that are extracted from GroZi-120 [22] dataset The1t2raining dataset contains of 676 training images representing 120 grocery products.
[Zhang et al., 2015]	Grocery Products	They have used a small database that contains of 3882 training images and 333 test images for 41 commercial product

		categories.

In this study a part of data is collected from DeepFashion dataset, a intimately available dataset with colorful apparel images. There's a aggregate of seven classes for discovery using Yolo v3. The classes taken from this dataset are Dress, Tee, Films, Jeans. For the remaining classes of Shirts, Caps and. Specs data is collected from colorful sources on the Internet and images collected from a mobile phone. Some of the data is also taken from vids of arbitrary places with people walking around. Images of different kind are taken i.e. solid background images substantially taken in workrooms like prints taken outside with various backgrounds like and arbitrary prints taken from mobile in our lot like. A aggregate of 1800 images are used for discovery out of which 80 i.e. 1440 images are used for training YOLO v3 and remaining 360 images are used for confirmation. We're using Residual Networks for color bracket. A aggregate of six classes are taken for this task. The six colors are-Black, Blue, Gray, Green, Pink, Red. For this task images are taken from Colorful fashion stores and also from my mobile phone by taking film land in AIT as well as Thammasat University. A aggregate of 6000 images are collected, 1000 per each class. These images are divided into 80 and 20 percent independently for training and confirmation sets. Fresh 100 images per each class, i.e. 600 images are collected for making a test set. For better performance these apparel images are cropped in such a way that only a patch of color images is used in training and testing process.

Discussion

From the above survey, we determined some open problems in product recognition and presented some proposals for researchers. First, there is no unification model to recognize all kinds of products such as garments, grocery products, and vehicles. We propose the use of the R-CNN [15] model to present product detection. R-CNN has achieved high performance on object detection and also it uses linear SVMs to classify each region [15]. It can be used to classify the product kind. Second, there is no specific model for recognizing a product on YouTube video or TV. The video consists of a number of frames and each frame is an image that consists of pixels. So, we treat the frame as we treat the image. Then, we can use R-CNN to detect and recognize the product in the frame.

Conclusion

Finally, this paper tries to cover the most important product recognition approaches and identifies state-of-the-art product recognition. We have shown there are three subfields of product recognition, which are product recognition on a mobile device, clothing recognition, and grocery product recognition. We have presented different methods for each subfield, discussed each method, and shown its strengths and weaknesses. We have presented datasets that have been used in product recognition, explaining each dataset and showing their categories. In the end, we showed some open problems in product recognition and presented some pro- proposals for researchers.

References

[1]P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33[5]:898–916, 2011.

[2]V. Chandrasekhar, G. Takacs, D. Chen, S. Tsai,R. Grzeszczuk, and B. Girod. Chog: Compressed his- program of gradients a low bit-rate feature descriptor. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2504–2511. IEEE, 2009.

[3]H. Chen, A. Gallagher, and B. Girod. Describing clothing by semantic attributes. In *Computer Vision–ECCV 2012*, pages 609–623. Springer, 2012.

[4]T. Chen, K.-H. Yap, and D. Zhang. Discriminative bag-of-visual phrase learning for landmark recognition. In *Acoustics, Speech and Signal Processing [ICASSP], 2012 IEEE International Conference on*, pages 893–896. IEEE, 2012.

[5]M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, and S. Hu. Global contrast-based salient

region detection.*Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 37[3]:569–582, 2015.

[6]N. Dalal and B. Triggs.Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition [CVPR'05]*, volume 1, pages 886–893. IEEE, 2005.

[7]J. Deng, A. Berg, S. Satheesh, H. Su, A. Khosla, and L. Fei- Fei. Imagenet large scale visual recognition competition 2012 [ilsvrc2012], 2012.

[8]J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei- Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009.

[9]M. Eichner, M. Marin-Jimenez, A. Zisserman, and V. Ferrari. Articulated human pose estimation and search in [almost] unconstrained still images. *ETH Zurich, D-ITET, BIWI, Technical Report No*, 272, 2010.

[10]M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge. *International journal of computer vision*, 88[2]:303– 338, 2010.

[11]P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ra- manan. Object detection with discriminatively trained part-based models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32[9]:1627–1645, 2010.

[12]P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph- based image segmentation. *International Journal of Com- puter Vision*, 59[2]:167–181, 2004.

[13]M. George and C. Floerkemeier. Recognizing products: A per-exemplar multi-label image classification approach. In *Computer Vision–ECCV 2014*, pages 440–455. Springer, 2014.

[14]B. Girod, V. Chandrasekhar, D. M. Chen, N.-M. Che- ung, R. Grzeszczuk, Y. Reznik, G. Takacs, S. S. Tsai, and R. Vedantham. Mobile visual search. *Signal Processing Magazine, IEEE*, 28[4]:61–76, 2011.

[15]R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Computer Vision and Pattern Recognition [CVPR], 2014 IEEE Conference on*, pages 580–587. IEEE, 2014.

[16]J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In *Advances in neural information processing systems*, pages 545–552, 2006.

[17]J. He, J. Feng, X. Liu, T. Cheng, T.-H. Lin, H. Chung, and S.-F. Chang. Mobile product search with bag of hash bits and boundary reranking. In *Computer Vision and Pattern Recognition [CVPR], 2012 IEEE Conference on*, pages 3005–3012. IEEE, 2012.

[18]H. Jegou, M. Douze, and C. Schmid. Product quantization for nearest neighbor search. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33[1]:117–128, 2011.

[19]Y. Kalantidis, L. Kennedy, and L.-J. Li. Getting the look: clothing recognition and segmentation for automatic product suggestions in everyday photos. In *Proceedings of the 3rd ACM conference on International conference on multimedia retrieval*, pages 105–112. ACM, 2013.

[20]J. Kim, C. Liu, F. Sha, and K. Grauman. Deformable spatial pyramid matching for fast dense correspondences. In *Com- puter Vision and Pattern Recognition [CVPR], 2013 IEEE Conference on*, pages 2307–2314. IEEE, 2013.

[21]A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012.

[22]D. G. Lowe. Distinctive image features from scale- invariant keypoints. International journal of computer vi- sion, 60(2):91–110, 2004.

[23]D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In Computer Vision and Pattern Recogni- tion, 2006 IEEE Computer Society Conference on, volume 2, pages 2161–2168. IEEE, 2006.

[24]C. Rother, V. Kolmogorov, and A. Blake. Interactive fore- ground extraction using iterated graph cuts, 2004. SIG- GRAPH.

[25]C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interac- tive foreground extraction using iterated graph cuts. ACM Transactions on Graphics (TOG), 23(3):309–314, 2004.

[26]C. Silpa-Anan and R. Hartley. Optimised kd-trees for fast image descriptor matching. In Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, pages 1–8. IEEE, 2008.

[27]S. S. Tsai, D. Chen, V. Chandrasekhar, G. Takacs, N.-M. Cheung, R. Vedantham, R. Grzeszczuk, and B. Girod. Mo- bile product recognition. In Proceedings of the international conference on Multimedia, pages 1587–1590. ACM, 2010.

[28]S. S. Tsai, D. Chen, J. P. Singh, and B. Girod. Rate-efficient, real-time cd cover recognition on a camera-phone. In Pro- ceedings of the 16th ACM international conference on Mul- timedia, pages 1023–1024. ACM, 2008.

[29]S. S. Tsai, D. Chen, G. Takacs, V. Chandrasekhar, J. P. Singh, and B. Girod. Location coding for mobile image retrieval. In Proceedings of the 5th International ICST Mobile Multime- dia Communications Conference, page 8. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunica- tions Engineering), 2009.

[30]P. Viola and M. Jones. Robust real-time object detection.International Journal of Computer Vision, 4:51–52, 2001.

[31]K. Yamaguchi, M. H. Kiapour, L. E. Ortiz, and T. L. Berg.Parsing clothing in fashion photographs. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pages 3570–3577. IEEE, 2012.

[32]Y. Yang and D. Ramanan. Articulated pose estimation with flexible mixtures-of-parts. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pages 1385–1392. IEEE, 2011.