



HAL
open science

Bag-of-word based brand recognition using Markov Clustering Algorithm for codebook generation

Yannick Benezeth, Aurélie Bertaux, Aldric Manceau

► **To cite this version:**

Yannick Benezeth, Aurélie Bertaux, Aldric Manceau. Bag-of-word based brand recognition using Markov Clustering Algorithm for codebook generation. IEEE International Conference on Image Processing (ICIP), 2015, Québec, France. hal-01177779

HAL Id: hal-01177779

<https://u-bourgogne.hal.science/hal-01177779>

Submitted on 17 Jul 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Bag-of-word based brand recognition using Markov Clustering Algorithm for codebook generation

Yannick Benezeth, Aurélie Bertaux, Aldric Manceau
LE2I, Université de Bourgogne
BP 47870, 21078 Dijon Cedex, France
{yannick.benezeth,aurelie.beraux,aldric.manceau,}@u-bourgogne.fr

Abstract—In order to address the issue of counterfeiting online, it is necessary to use automatic tools that analyze the large amount of information available over the Internet. Analysis methods that extract information about the content of the images are very promising for this purpose. In this paper, a method that automatically extract the brand of objects in images is proposed. The method does not explicitly search for text or logos. This information is implicitly included in the *Bag-of-Words* representation. In the *Bag-of-Words* paradigm, visual features are clustered to create the visual words. Despite its shortcomings, *k-means* is the most widely used algorithm. With *k-means*, the selection of the number of visual words is critical. In this paper, another clustering algorithm is proposed. *Markov Cluster Algorithm (MCL)* is very fast, does not require an arbitrary selection of the number of classes and does not rely on random initialization. First, we demonstrate in this paper that *MCL* is competitive to *k-means* with a number of cluster experimentally selected. Second, we show that it is possible to identify brand from objects in images without previous knowledge about visual identity of these brands.

I. INTRODUCTION

Counterfeits are illegally made products that resemble the authentic goods but are typically of lower quality [20]. The global market for counterfeits accounts for approximately 7% of world trade, representing hundreds of billions of dollars. The negative impact that counterfeiting can have on brands is the erosion of brand equity that could occur if consumers are aware that some portion of the available stock of a brand is actually counterfeit. This situation could erode confidence in the brand and reduce the status value that is sometimes associated with brand ownership [9]. As the existence of counterfeits is detrimental to the affected brands, the need to determine the distributors of counterfeits is essential.

Nowadays, most transactions are conducted online. This also applies to counterfeiters. Websites selling counterfeit goods, including luxury goods, generate more than 92 million visits per year [10]. Examining online traffic patterns is vital in identifying counterfeiter websites. In order to analyze the large amount of information available over the Internet, it is necessary to use automatic tools. Web crawlers browse the Internet and extract features about the server hosting the web page (URL, location, etc.) and the web page content. The text is obviously the easiest part of the web page that can be analyzed but images are also very critical.

Large amount of researches on content-based image retrieval and object recognition have been carried out in the



Fig. 1. Illustration of the strong visual identity of two well-known brands. In the first line, Vuitton's bags are presented and Chanel's ones in the second line.

past two decades. However, very few papers focus on the automatic recognition of product brands. The most related works use text detection and Optical Character Recognition (*OCR*) algorithms to extract the name of the brand in the image. Word detection and recognition in uncontrolled images is a challenging problem but many successful methods have been proposed recently [19], [5]. Another possible strategy is to extract logos in the image. However, logos often appear superimposed on objects of any geometry. Early work on logo detection and recognition was concerned with providing some automatic support to the logo registration process [11]. More recently, methods based on the Bag-of-Words (*BoW*) paradigm have showed good capability to recognize logo in real world images [16], [1]. To the best of our knowledge, there is no existing work on brand recognition of a product in an image that does not explicitly search for text or logos. However, some brands have a strong visual identity and propose instantly recognizable objects. In Fig 1, six bags of two well-known brands are presented. Both brands are easily recognizable by a person interested in fashion. From these observations, we propose in this paper a method that automatically extract the brand of objects in images without explicitly describing the visual identity of the brand. This information is implicitly included in the *BoW* representation.

Our method is based on the successful *BoW* paradigm. This idea originated for text categorization, where a codebook is

created from words contained in the text. Then, classification is possible by observing the frequency of each word in the text. The visual counterpart and its extensions has been successfully applied to object recognition in images (e.g. in [17], [6], [12]). Images are represented by histograms where each element represents the frequency of each visual word in the image. Usually, the *BoW* image representation involves three steps: features detection, features description, and codebook generation. Despite its limitations, *k-means* is the most widely used algorithm to construct the visual dictionary (or codebook). The objective is to cluster visual features into k visual words based on their proximity in the feature space. The limitations of *k-means* concern the critical selection of the number of classes and the random initialization. It has been reported that performance improves steadily as the codebook grows but large codebooks may encounter the over-fitting problems [7]. Then, multiple repetitions of *k-means* clustering on the same data may produce different visual words. In this paper, we propose to use *Markov Clustering Algorithm (MCL)* for the codebook generation. *MCL* is capable to get through those two limitations.

In the following, an overview of the method is presented in section II. The section III introduces the *MCL* clustering algorithm and section IV presents a comparison between *MCL* and *k-means* and results on brand recognition in images.

II. OVERVIEW OF THE APPROACH

The goal of this work is to recognize the brand of objects in images. We focus our efforts on brands with a strong visual identity. This visual identity is a key element of marketing communications. Consistent use of logos, specific color palettes and other characteristics help consumers to easily identify the brand. Fig. 1 illustrates this concept with six bags from two brands with a strong visual singularity. In this work, we do not explicitly include the elements that define the visual identity of a brand but these elements are implicitly included in the *BoW* representation. Fig. 2 illustrates the *BoW* representation of two bags. These bags are characterized by a high frequency of visual words specific to these brands. We experimentally observe that these discriminating visual words often represent a portion of the logo. It is quite common that brands use variations of their logo on their product. But if the dictionary is discriminant enough, there are visual words that are shared by most objects of the brand.

BoW methods first construct a visual dictionary (codebook) from a task-dependent images dataset and then describe other images according to this codebook. At the end, each image is represented by a frequency of each visual word in the image. We choose David Lowe’s SIFT [13] as the interest point detector and descriptor. While there are other detectors and descriptors (such as MSER [14] or SURF [2] to name a few), SIFT has been extensively compared with other descriptors and it has been demonstrated that it performs very well in image matching task [15].

The SIFT detector, like most local feature detectors, results in a large number of interest points. In the *BoW* paradigm [17],

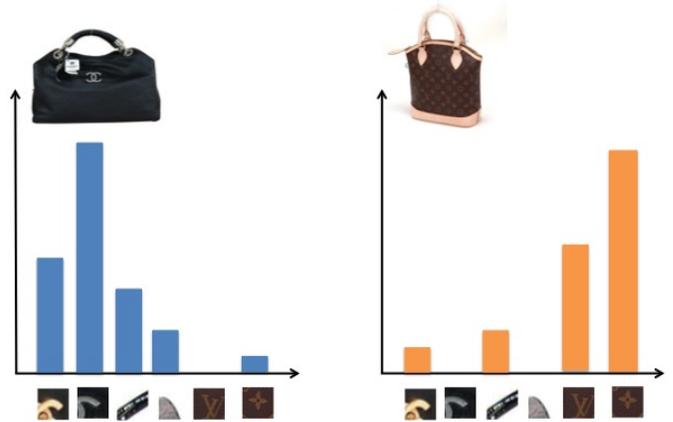


Fig. 2. Bag-of-Words representation of two bags. These bags are characterized by a high frequency of visual words specific to these brands.

these descriptors are summarized by quantizing and aggregating the features. We apply Markov Clustering Algorithm (*MCL*) to a large number of SIFT features to create the dictionary of visual words. This codebook is then used to quantize the extracted features by simply assigning the label of the closest cluster centroid. The final representation for an image is the frequency counts of the labeled SIFT features. This *BoW* representation of an image is later used in a supervised classification framework to determine which class to assign to the image.

III. CODEBOOK GENERATION WITH MARKOV CLUSTERING ALGORITHM

A. *MCL* algorithm

In this paper, we consider an unsupervised and efficient graph-mining algorithm for clustering: the *Markov Clustering Algorithm (MCL)* developed by S. Van Dongen in his thesis [18]. Indeed, this algorithm is well known in graph community due to the strong underlying mathematical theory, its scalability possibilities, easy implementation and good performance.

This algorithm finds cluster structure in graphs by a mathematical bootstrapping procedure. For that, it uses random walks through the graph and two operators. The mathematical concept of random walks is captured by the use of stochastic¹ (i.e. Markov) matrices. The first operator, called *expansion*, coincides with taking the power of a stochastic matrix using the regular matrix products. The second operator, called *inflation*, allows to keep the matrix stochastic and changes the probabilities by favoring more probable walks over less probable walks. In fact, this operator influences the granularity and hence the size of clusters.

A matrix is created from input data (an adjacency matrix if the input matrix is a graph). It is important to highlight that the input matrix has to be a square matrix. First, because there exists a strong mathematical effect that odd powers of

¹A column stochastic matrix is a non-negative matrix with the property that each of its columns sums to 1.

expansion obtain their mass from simple paths of odd length, and likewise for even, so a self loop to each node is added (*i.e.* add diagonal elements to the matrix). Then, the algorithm starts and iterates *expansion* and *inflation* until the matrix is steady. The algorithm starts to converge after 3-10 iterations. Even if the convergence is hard to prove, it always does if the input graph is symmetric [18]. Algorithm 1 presents the detailed procedure. The expansion step is shown in line 3, then lines 5 to 7 inflate the matrix and lines 8 to 10 convert the matrix to a stochastic matrix.

One of the first advantage of *MCL* is that the number of clusters are not defined by the user like many clustering methods (such as *k-means*). *MCL* is likewise scalable and able to manage a huge volume of data. It can also ignore outliers in most of case [3]. At last, this algorithm is very easy to implement. However, the main drawback of this algorithm is its high cost of computation [4]. *MCL* has been modified to be effective for the codebook generation in the *BoW* paradigm.

B. *MCL* for codebook generation

MCL algorithm is a good candidate to create the codebook in *BoW* methods. With this clustering algorithm, the number of clusters is not previously defined and is only determined by the dataset.

In the previous section, it is specified that the input matrix must be a square matrix. Each connection between nodes represents a link. To model this relationship, the euclidean distance between each visual features is computed. This matrix is the adjacency matrix. In the classical *MCL*, the input matrix is a binary matrix. For the codebook generation, the expansion step has been removed because the power operator get some negative influence to the result. The use of the expansion operator on non-binary matrix implies a meaningless single cluster output. The only parameter to adjust is the inflation parameter (we experimentally observed that 2 is a good choice). Without the expansion step, the *MCL* clustering is very fast.

IV. EXPERIMENTAL RESULTS

In this section, we compare the *MCL* clustering algorithm to the well-known *k-means*. We also demonstrate that the elements defining the visual identity of a brand are implicitly included in the *BoW* representation and can be used to automatically identify objects of a specific brand.

A. Experiment 1: Object recognition

In this first experiment, we established a dataset composed of three kinds of objects, *i.e.* bags, fragrances and watches. An illustration of this dataset is presented in Fig. 3. We collected about 700 images per class for this first experiment (mostly from Flickr website). Each set has been divided into training/test subsets. We used 100 images in the test subset and the rest is used for the training. The codebook has been generated using randomly selected images from the three training subsets. We have followed the methodology of the Classification task of Pascal VOC Challenge [8]. The objective

Require: A weighted undirected graph $G=(V,E)$, expansion parameter e , inflation parameter r

Ensure: A partitioning of V into disjoint components

```

1:  $M \leftarrow M(G)$  ;
2: while  $M$  is not a fixpoint do
3:    $M \leftarrow M^e$ ;
4:   for all  $i \in V$  do
5:     for all  $j \in V$  do
6:        $M[i][j] \leftarrow M[i][j]^r$ ;
7:     end for
8:     for all  $j \in V$  do
9:        $M[i][j] \leftarrow \frac{M[i][j]}{\sum_{k \in V} M[i][k]}$ ;
10:    end for
11:  end for
12: end while
13:  $H \leftarrow$  graph induced by non-zero entries of  $M$ 
14:  $C \leftarrow$  clustering induced by connected components of  $H$ 

```

Algorithm 1: *MCL* algorithm

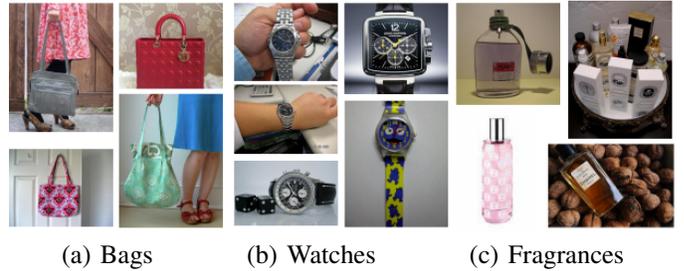


Fig. 3. Sample images from the dataset of *exp1*: object recognition.

of this experiment is to predict, for each of the three object classes, the presence/absence of at least one object of that class in a test image. Results are presented using the Average Precision (AP) as the area under the precision/recall curve. The visual codebook has been generated using *k-means* and *MCL* for the clustering step.

	Bag	Watch	Fragrance
k-means	84.53%	90.69%	88.26%
MCL	86.77%	88.16%	85.22%

TABLE I
AVERAGE PRECISION FOR THE CLASSIFICATION TASK USING THREE OBJECT CLASSES.

Results in Table I show that *MCL* is competitive compared to *k-means*. It should be noted that for *k-means*, the number of classes has been optimized with a greedy approach. Average precision is slightly better with the *Bag* class using *MCL* and a little below with the *Watch* and *Fragrance* classes.

B. Experiment 2: Brand recognition

In this second experiment, we show that it is possible to automatically extract the brand of objects in images without explicitly describing the visual identity of the brand, *i.e.* without explicitly looking for a logo or a color palette. We established a dataset composed of images of bags of three



Fig. 4. Sample images from the dataset of *exp2*: brand recognition.

brands, *i.e.* Vuitton, Guess and Chanel. An illustration of this dataset is presented in Fig. 4. We collected about 160 images per class for this second experiment. Each set has been divided into training/test subsets. We used 100 images in the training subset and the rest is used for the tests. The objective of this experiment is to identify the brand of a bag given an image. Results are presented using the Average Precision (AP). The visual codebook has been generated using *k-means* and *MCL* for the clustering step.

	Chanel	Vuitton	Guess
k-means	96.03%	99.26 %	97.23%
MCL	97.00%	99.08%	97.22%

TABLE II

AVERAGE PRECISION FOR THE BRAND RECOGNITION TASK USING IMAGES OF BAGS.

First, it is possible to observe in Table II that results are surprisingly very high, above 95%, validating our approach. An empirical analysis of the most discriminant visual words showed that the most important ones were located on logos. These logos appear as a texture for the bags of *Guess* and *Vuitton*. Second, as for *experiment 1*, we obtained comparable results than with *k-means* and *MCL*.

Obviously, some brands have a visual identity more discreet and more commonplace and we cannot expect such results with all brands. In some cases, other regular image analysis methods can be more useful (for example *OCR*). For some brands, visual identity is mainly defined by a specific color palette. In these cases, it is possible to include other relevant feature to capture the visual identity of the brand.

V. CONCLUSION

The first contribution presented in this paper is an automatic image analysis tool that can extract the brand of an object in an image. While other methods in the state of the art use text detection and *OCR* to extract the name of the brand in the image or try to detect the logo, in this work, we do not explicitly include the elements that define the visual identity of a brand. These elements are implicitly included in the *BoW* representation. The second contribution is the use of *Markov Clustering Algorithm* for the codebook generation. This algorithm has some advantages compared to *k-means*, namely it does not require a selection of the number of visual words and it does not rely on random initialization.

In this paper, we experimentally compare *MCL* and *k-means* on a classical object recognition task. Then, we demonstrate

the capability to recognize the brand of objects in images using the *BoW* paradigm.

Future works will focus on a rigorous validation of *MCL* on widely used datasets (such as Pascal VOC Challenge or Image Net). We will also try to apply the proposed brand recognition methodology on images from websites selling counterfeit goods to validate our approach.

REFERENCES

- [1] BAGDANOV, A., BALLAN, L., BERTINI, M., AND DEL BIMBO, A. Trademark matching and retrieval in sports video databases. In *Proc. of the int. workshop on multimedia information retrieval* (2007), pp. 79–86.
- [2] BAY, H., TUYTELAARS, T., AND VAN GOOL, L. Surf: Speeded up robust features. In *European Conference on Computer Vision*. 2006, pp. 404–417.
- [3] BROHEE, S., AND VAN HELDEN, J. Evaluation of clustering algorithms for protein-protein interaction networks. *BMC bioinformatics* 7, 1 (2006), 488.
- [4] CHAKRABARTI, D., AND FALOUTSOS, C. Graph mining: Laws, generators, and algorithms. *ACM Computing Surveys* 38, 1 (2006).
- [5] COATES, A., CARPENTER, B., CASE, C., SATHEESH, S., SURESH, B., WANG, T., WU, D. J., AND NG, A. Y. Text detection and character recognition in scene images with unsupervised feature learning. In *Proc. of the Int. Conf. on Document Analysis and Recognition* (2011), pp. 440–445.
- [6] CSURKA, G., DANCE, C., BRAY, C., FAN, L., , AND WILLAMOWSKI, J. Visual categorization with bags of keypoints. In *Pattern Recognition and Machine Learning in Computer Vision Workshop* (2004).
- [7] CSURKA, G., DANCE, C., FAN, L., WILLAMOWSKI, J., AND BRAY, C. Visual categorization with bags of keypoints. In *Workshop on statistical learning in computer vision, ECCV* (2004), vol. 1, pp. 1–2.
- [8] EVERINGHAM, M., VAN GOOL, L., WILLIAMS, C. K. I., WINN, J., AND ZISSERMAN, A. The PASCAL Visual Object Classes Challenge 2011 (VOC2011) Results. <http://www.pascal-network.org/challenges/VOC/voc2011/workshop/index.html>.
- [9] GREEN, R., AND SMITH, T. Countering brand counterfeiters. *Journal of International Marketing* 10, 4 (2002), 89–106.
- [10] HARLAND, B. Report sheds light on scale and complexity of online piracy and counterfeiting problem, January 2011. Mark Monitor press release.
- [11] KIM, Y. S., AND KIM, W. Y. Content-based trademark retrieval system using visually salient feature. In *Proc. of int. conf. on Computer Vision and Pattern Recognition* (1997), p. 307312.
- [12] LOPEZ-SASTRE, R. J., TUYTELAARS, T., ACEVEDO-RODRIGUEZ, F. J., AND MALDONADO-BASCO N, S. Towards a more discriminative and semantic visual vocabulary. *Computer Vision and Image Understanding* 115 (2011), 415–425.
- [13] LOWE, D. Object recognition from local scale-invariant features. In *International Conference on Computer Vision* (1999), vol. 2, pp. 1150–1157.
- [14] MATAS, J., CHUM, O., URBAN, M., AND PAJDLA, T. Robust wide-baseline stereo from maximally stable extremal regions. *Image and vision computing* 22, 10 (2004), 761–767.
- [15] MIKOLAJCZYK, K., AND SCHMID, C. A performance evaluation of local descriptors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 27, 10 (2005), 1615–1630.
- [16] SAHBI, H., BALLAN, L., SERRA, G., AND DEL BIMBO, A. Context-dependent logo matching and recognition. *IEEE Transactions on Image Processing* 22, 3 (2013), 1018–1031.
- [17] SIVIC, J., AND ZISSERMAN, A. Video google: A text retrieval approach to object matching in videos. In *Proc. of the International Conference on Computer Vision* (2003).
- [18] VAN DONGEN, S. Graph clustering by flow simulation. <http://www.library.uu.nl/digiarchieff/dip/diss/1895620/inhoud.htm>.
- [19] WANG, K., BABENKO, B., AND BELONGIE, S. End-to-end scene text recognition. In *International Conference on Computer Vision* (2011), pp. 1457–1464.
- [20] WILCOX, K., HYEONG, M., AND SANKAR, S. Why do consumers buy counterfeit luxury brands? *Journal of Marketing Research* 46, 2 (2009), 247–259.