Content Based Image Retrieval and Its Application to Product Recognition

Petros Fotios Alvanitopoulos, Andrian Moroi, George Bagropoulos, Kieran Dundon

HAL Id: hal-01385341
https://inria.hal.science/hal-01385341

Submitted on 21 Oct 2016

Content Based Image Retrieval and its Application to Product Recognition

Petros Fotios Alvanitopoulos[1], Andrian Moroi[1], George Bagropoulos[1], Kieran Dundon[1]

[1]ChannelSight,
Research and Development Center,
Suite 3, Anglesea House, Carysfort Ave, Blackrock,
Co. Dublin, Ireland.

{petros,andrian, george, kieran}@channelsight.com

Abstract. Product Recognition is a challenging problem in many practical applications. This paper presents a new approach for product recognition. By utilizing a set of crawlers our task is to extract informative content from web pages and automatically recognize products found on web pages. A set of images is extracted from each web page and then a new "content-based" image retrieval technique is performed to rank the images from our product catalog. The proposed content-based image retrieval technique utilizes the Empirical Mode Decomposition and processes the first extracted component of the source image. This component maintains the highest local spatial variations of the source image. An adaptive local-threshold technique is applied for the extraction of edges. A quantized and normalized histogram is created for the representation of images. Simulation results reveal that the proposed method is a promising tool for the challenge task of product recognition.

1   Introduction

Product Recognition (PR) is a challenging task. To date, an increasing number of internet users buy products from well-known e-commerce web sites such as Amazon, Newegg and Walmart. Moreover, the interest of manufacturers for product coverage in web sites has grown in recent years. Many companies invest in information technology in order to increase their sales by means of implementing efficient product search services. One of the most popular approaches is to offer to clients straightforward web-services. These services are dedicated to find reviews, specifications availability and prices of products from different retailers with the ultimate goal to purchase them.

Today the tremendously evolution in portable electronic devices like smartphones and tablets, provides a perfect environment where more and more users tend to shop

online. Using these devices several web-services are implemented, having as goal to provide an attractive and alternative environment to proceed to shopping actions. Current product-image search systems on the market include Google Goggles [1], Vuforia [2] and Flow [3].

Product Image Retrieval (PIR) is an emerging field of content-based image retrieval (CBIR). Several algorithms and systems have been introduced in literature for this new research topic [4-10]. From these studies it can be seen that the product recognition problem falls into two categories: a) systems where users get online information regarding a product by snapping a picture of a product with a camera-device (smartphone or tablet) and b) systems where a vertical search is performed based on domain-specific search such as "Philips TVs" domain. In the latter case, focused crawlers/parsers extract the informative content from web product-pages and try to recognize the product by taking into consideration text content and corresponding product images that appear in the web product page [10]. Almost all the existing PR approaches involve CBIR techniques. A wide variety of CBIR have already been proposed in the literature (e.g. [11-16]).

In this work a new CBIR method is proposed for the PR challenging task. Considering the scenario that customer seeks for a product, our goal is to create a search engine that recognizes products from websites with high commercial transactions. Images and description of products constitute the informative content of webpages. The downloaded images from retailers' web pages are represented by a vector of visual features. The proposed CBIR algorithm retrieves similar images from a database which includes images for several products from a large set of manufacturers.

The proposed CBIR method utilizes the Empirical Mode Decomposition (EMD) in order to extract the edges of images. The highest local spatial variations are localized with the representation of images in the Hue-Saturation-Value (HSV) color space. EMD is an adaptive filtering method introduced by Huang and et al. [17] in 1998. It does not require any a priori assumption regarding the basis and it is based on local characteristics of the signal (extrema). Some modifications of the EMD applied for Image Processing can be found in [18-20]. After the extraction of the first component a local-threshold adaptive approach [21] is performed in order to extract qualitative edges and interested points of images. Then a quantized color histogram is created. Twenty five (25) perceptual colors are selected based on the color image quantization method proposed in [22].

The rest of this paper is organized as follows. In Section 2, an overview of the system architecture is introduced. The proposed CBIR is presented in Section 3. In Section 4, performance results are shown. Section 5 concludes the paper.

2 System Architecture

2.1 Overview of the System

The World Wide Web (www) has become an important data resource. The rapid growth of its size reveals the need for effective web search engines. However, the task

of extracting information from www is non-trivial. Conventional techniques for extracting information include web crawlers.

Crawler is a program that downloads web pages and extracts raw information regarding the downloaded web pages [10]. In general web crawler is a keystone in many web services (e.g. shopping search engines, Internet Archives). In our system a PR search engine is implemented taking into consideration the vertical search approach. In the vertical search scenario focused web crawlers are utilized optimized for specific domain. Focused web crawlers which seek and maintain pages related to a specific topic, provide a narrow segment of the www, and thus they constitute a robust and effective component for the design of a potential high-performance PR system.

Since our goal is the PR challenging task, it is clear that an efficient approach requires the integration of both textual and visual content. A brief description of the overall workflow of our system is shown in Fig. 1. From Fig. 1 it can be seen that the PR system incorporates four major components: a) the Scheduler, b) the Focused Crawler, c) the HTML-Parser, d) the Search UI component and e) the PR Unit.
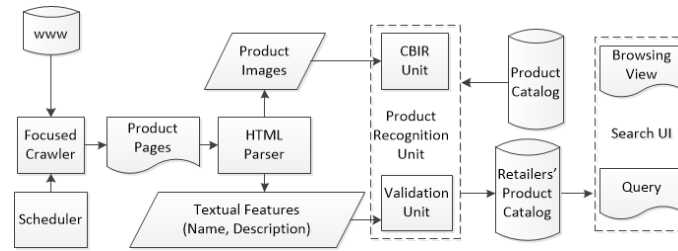

Fig. 1. Workflow of the PR system.

The Scheduler is responsible to create crawl sessions devoted to particular websites. The Scheduler triggers Focused Crawlers to navigate to websites and extract the product pages from online retailers. The HTML-Parser processes the source code from a product page and selects the informative content including product title, price, description, availability, metadata and product images. The Validation Unit verifies the matching decision produced from the CBIR algorithm between a product page and a product that appears in the Product Catalog database through a text analysis process. The Search UI component is the graphical user interface. It includes features such as product querying, product purchasing, product description, and list of prices regarding an item found in different retailers.

The PR unit is the most intelligent part of the system. This component incorporates the proposed CBIR technique and analyzes images in order to represent them in an appropriate feature space where the similarity between images is more effective. In general CBIR process implies several challenging issues, including analysis of low-level features (e.g. color, shape, texture, spatial layout, etc.) and creation of feature vectors. The size of the feature vector affects the quality and the speed of a CBIR system. Basically, the key to attain efficient CBIR is to compromise these factors.

Thus, a real-time and accurate CBIR system, for the representation of images, requires a small size of feature vector, while at the same time it has to maintain its accuracy through a large set of features.

3 Proposed Method for PIR

Let C be the color product image with width w and height h. Considering the HSV color space the Value channel V of the color image C is selected as the source image for the edge detection method. The Value component of the HSV color space represents the brightness of an image and it's vital to the definition of edge points. The reason for the selection of the HSV color space for the edge detection process is the fact that HSV corresponds similar to the way humans perceive color [23] and it is well-known, that in comparison to RGB, it is more accurate in shadow distinction. However, it should be mentioned that the selection of the HSV color is not panacea and different color spaces may also be suggested in this step. For the edge detection process a preprocessing analysis of the grayscale image V is performed by utilizing the EMD to filter the image and extract a more informative component. The application of the EMD for the definition of edge points is described below.

3.1 Bidimensional Empirical Mode Decomposition BEMD

EMD is a proper signal processing technique for analyzing nonlinear and non-stationary data. It is an adaptive technique and the extraction of the basis is attained based on local characteristics (extrema) of the analyzed signal. Bearing in mind the Hilbert transform and the concept of instantaneous frequency the cooperation of these definitions with the EMD forms a powerful tool for time-frequency analysis, named Hilbert-Huang-Transform (HHT) [17]. The EMD decomposes a signal into its intrinsic mode functions (IMFs). These components exhibit well behavior under the Hilbert transform.

In recent years, the EMD was applied to analyze also two-dimensional data such as images. Proposed versions of the BEMD can be found in the literature [18-20]. In the proposed CBIR algorithm a modification of Window Empirical Mode Decomposition (WEMD) [18] is used in combination with the improved bidimensional EMD (IEMD) [19]. The proposed approach of the BEMD is given below:

1. Initialization. Set $r_0=V$, $i=1$ where V is the grayscale image from the Value channel of the color product image.
2. Extract all extrema points (maxima-minima) in r_{i-1}
3. Calculate the mean signal m_{i-1} described as follows:
   (a) Initialization: Set the maximal size N of the window and set its original size M to 3.
   (b) If the number of maxima is equal with the number of minima in the window M, then compute the mean value from all pixels appeared in the window of size M, and turn to step (d);

(c) Set $M=M+2$. If $M<N$, turn to step (b), otherwise compute the mean value from all pixels appeared in the window of size N;

(d) Turn to the next pixel, set $M=3$, turn to step (b), repeat until all pixels of the image are processed, then compute the mean surface m_{i-1};

4. Compute $imf_{i-1}=r_{i-1}-m_{i-1}$, $r_i = m_{i-1}$, $i=i+1$;

5. Repeat steps (2)-(5) until $i=n$ where n equals to the number of IMFs.

The summation of all the IMFs and the residue image R can reconstruct the original grayscale image V.

$$V = \sum_{I=1}^{n} IMF_i + R \tag{1}$$

The maximum size N of the window can be determined based on the maxima and minima maps created from the source image. The Euclidean distance to the nearest other local maximum (minimum) point is estimated and denoted as D_i (D_j) and the maximum size N of the window is computed by the following equation:

$$N = \max\{\min\{D_i\}, \min\{D_j\}\}. \tag{2}$$

where min/max{} denotes the minimum/maximum value of the elements in the array{}.

The extrema points are defined based on the method proposed in [19]. According to this method five different types of extrema are proposed. Assuming a square mask window S of odd size, the center pixel $S_{i,j}$ is an extrema if one of the following conditions are fulfilled:

- First type of extrema:
  If $S_{i,j}=S_{i-1,j}=S_{i+1,j}$ and

$$\bigcap_{l=i-1}^{i+1} \bigcap_{q=\pm 1} S_{l,j} > S_{l,j+q} \tag{3}$$

  $S_{i,j}$ is a local maximum.
  If $S_{i,j}=S_{i-1,j}=S_{i+1,j}$ and

$$\bigcap_{l=i-1}^{i+1} \bigcap_{q=\pm 1} S_{l,j} < S_{l,j+q} \tag{4}$$

$S_{i,j}$ is a local minimum.

- Second type of extrema:
  If $S_{i,j}=S_{i,j-1}=S_{i,j+1}$ and

$$\bigcap_{l=j-1}^{j+1} \bigcap_{q=\pm 1} S_{i,l} > S_{i+q,l} \tag{5}$$

  $S_{i,j}$ is a local maximum.
  If $S_{i,j}=S_{i,j-1}=S_{i,j+1}$ and

$$\bigcap_{l=j-1}^{j+1} \bigcap_{q=\pm 1} S_{i,l} < S_{i+q,l} \tag{6}$$

  $S_{i,j}$ is a local minimum.

- Third type of extrema:

If $S_{i,j}=S_{i-1,j+1}=S_{i+1,j-1}$ and

$$\bigcap_{d=-1}^{+1}\left(\bigcap_{q=-1}^{+1}\bigcap_{\substack{l=-1\\l\neq-q}}^{+1} S_{i+d,j-d} > S_{i+d+q,j-d+l}\right) \tag{7}$$

$S_{i,j}$ is a local maximum.
If $S_{i,j}=S_{i-1,j+1}=S_{i+1,j-1}$ and

$$\bigcap_{d=-1}^{+1}\left(\bigcap_{q=-1}^{+1}\bigcap_{\substack{l=-1\\l\neq-q}}^{+1} S_{i+d,j-d} < S_{i+d+q,j-d+l}\right) \tag{8}$$

Si,j is a local minimum.

- Fourth type of extrema:

If $S_{i,j}=S_{i-1,j+1}=S_{i+1,j-1}$ and

$$\bigcap_{d=-1}^{+1}\left(\bigcap_{q=-1}^{+1}\bigcap_{\substack{l=-1\\l\neq q}}^{+1} S_{i+d,j+d} > S_{i+d+q,j+d+l}\right) \tag{9}$$

$S_{i,j}$ is a local maximum.
If $S_{i,j}=S_{i-1,j-1}=S_{i+1,j-1}$ and

$$\bigcap_{d=-1}^{+1}\left(\bigcap_{q=-1}^{+1}\bigcap_{\substack{l=-1\\l\neq q}}^{+1} S_{i+d,j+d} < S_{i+d+q,j+d+l}\right) \tag{10}$$

$S_{i,j}$ is a local minimum.

- Fifth type of extrema:

If

$$\bigcap_{l=-1}^{+1}\bigcap_{q=-1}^{+1} S_{l,j} > S_{i+l,j+q} \tag{11}$$

$S_{i,j}$ is a local maximum.
If

$$\bigcap_{l=-1}^{+1}\bigcap_{q=-1}^{+1} S_{l,j} < S_{i+l,j+q} \tag{12}$$

$S_{i,j}$ is a local minimum.

3.2   Edge detection points

In this section the edges of the product image are defined. Once the first IMF component (IMF1) is extracted, through the application of BEMD to the Value channel, a local edge detection technique is used. The applied edge detection algorithm is an adaptive technique introduced in [21]. This method deals with degradations appeared

due to shadows, non-uniform illumination, low contrast and high signal-dependent noise. Moreover, it runs automatically without requiring the adjustment of a set of design parameters in each execution. It is worthy to mention that this method belongs to local-threshold techniques. In comparison to global-threshold methods, these techniques provide better performance since they utilize the local area information in order to estimate a threshold value per pixel. Each pixel is classified based on its threshold value into foreground or background class. A flow chart with the major steps of the edge detection algorithm is presented in Fig. 2.
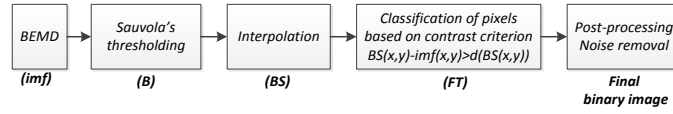

Fig. 2. Flow chart of the edge detection process.

In the first step of the original method a preprocessing filtering stage of the grayscale is required by utilizing a low-pass Wiener filter [24]. This filter is usually used for image restoration. However, in this paper this step is skipped since a preprocessing filtering is already achieved through the extraction of the first IMF component. The second step of the algorithm implements a rough classification of pixels into foreground and background clusters. This is attained via the Sauvola's approach [25]. The initial local threshold estimation is given by:

$$T = m + (1 - k\left(1 - \frac{s}{R}\right))$$

(13)

where m is the mean value of the rectangular window, T is the threshold value for the center pixel of the window and s is the variance value. R is a constant fixed to 128 and k is set to 0.2 according to [21]. Assuming the first IMF image imf the binary image B is extracted where 1s' represent the foreground pixels.

A background surface BS is then computed by following two rules: For zero pixels in image B the corresponding value at BS image equals to imf corresponding values. For the foreground pixels the corresponding values of the BS image are computed via interpolation described in the following equations:

$$BS(i,j) = imf(i,j) \; if \; B(i,j) = 0$$

(14)

otherwise

$$BS(i,j) = \frac{\sum_{ix=i-dx}^{i+dx} \sum_{iy=j-dy}^{j+dy} (imf(ix,iy)(1 - B(ix,iy)))}{\sum_{ix=i-dx}^{i+dx} \sum_{iy=j-dy}^{j+dy} (1 - B(ix,iy))}$$

(15)

The window size of dx x dy is set to 5x5.

The next step of the process involves the final thresholding phase, where the binary image FT is estimated taking into account the imf image and the background surface BS. The foreground pixels are determined based on the distance value between the imf image and the background surface BS. According to this method, a threshold d is

proposed and a foreground pixel is recognized if the calculated distance value exceeds the threshold d value. The final thresholding binary image at each point (x,y) is given by the following equations:

$$FT(x,y) = \begin{cases} 1 & if \ \ BS(x,y) - imf(x,y) > d\big(BS(x,y)\big) \\ 0 & otherwise. \end{cases} \qquad (16)$$

The threshold d for each pixel simulates a logistic sigmoid function and it is given by the following equation:

$$d\big(B(x,y)\big) = q\delta \left(\frac{(1-p_2)}{1 + exp\left(\frac{-4BS(x,y)}{b(1-p_1)} + \frac{2(1+p_1)}{(1-p_1)}\right)} + p_2 \right) \qquad (17)$$

where the variable b represents the average background value, the parameter q is equal to 0.6, the p_1 is set to 0.5 and p_2 to 0.8 according to [21] after extensive experimental work. The variable δ is the average distance between foreground and background values of the pixels and it is expressed by the following formula:

$$\delta = \frac{\sum_x \sum_y (BS(x,y) - imf(x,y))}{\sum_x \sum_y B(x,y)} \qquad (18)$$

The final step of the edge detection method includes the post-processing enhancement of the binary image by applying shrink and swells filters. These filters aim to remove isolated pixels and fill potential breaks and gaps. A detailed description of the algorithm can be found in [26].

3.3   Interested Points Detection

In the proposed method neurobiological findings are taken into consideration for the design of the PIR method. It is well-known that the ganglion cells (GC) constitute the unique output of the retina to the primary visual cortex. In this early neuronal vision component (retina) a preprocessing step is responsible to filter the initial visual information. This phase is required to minimize the amount of the transmitted information due to the limited bandwidth of the channel between retina and primary visual cortex. In retina the GCs and the connected receptive fields (RF) are the major components. The RFs of GCs have a center-surround mechanism. The RF consists of two concentric areas, the center and the surround that have antagonistic effects on the response of a GC. Consequently, this mechanism sends information regarding two different regions and the brain is responsible to run the "filling-in" mechanism in order to create the scene. Inspired by the biological visual system of mammals, only the interested points of images are taken into consideration for the next processing step. Thus, motivated by the biological vision systems, the considered interested points are the edges of images and the surrounded pixels of them. Below follows a detailed description of the selection of the interested points:

Let $L=\{P_1(x_1,y_1) , P_2(x_2,y_2), ..., P_n(x_n,y_n)\}$ be a set of n edge points extracted from the edge detection algorithm. Suppose that x_i, y_i are the x-coordinate and y-coordinate respectively of the ith point. Assume a window of constant size $k \ x \ k$. We slide the window to each edge point and seek for the interested points in the (k^2-1) neighbor points. For each edge point a unique window $k \ x \ k$ is associated containing k^2 poten-

tial interested points. It is reasonable to assume that the edge points belong to the set of the interested points. Thus, our investigation is located only to the rest (k^2-n) neighbor points. n equals to the number of edge points in the window. In our analysis the constant size k of the window is set to 3. The entropy is utilized as criterion for the selection of the rest interested points. Let $f_1, f_2, ..., f_m$ be the observed gray-level frequencies in the sliding window of the *imf* image. The percentage of occurrence of a specific gray level p_i equals to:

$$p_i = \frac{f_i}{k^2}, \qquad \sum_{i=1}^{m} f_i = k^2, \ i = 1,2,...,m. \tag{19}$$

The entropy H of a set of pixels appeared in the window is defined as:

$$H = -\sum_{i=1}^{s} p_i \log_2 p_i \tag{20}$$

where s represents the number of pixels in the set. For each gray level appeared at the window the algorithm removes the corresponding pixels and estimates the entropy of the subset that includes the remaining pixels. Our target is to find the gray level that produces the maximum entropy of the subset of pixels in the window and therefore to select only the subset of pixels that it is more heterogeneous.

3.4 Normalized Quantized Histogram

Color is one of the essential image indexing features in CBIR. The description of color information in images can be expressed by the means of color histograms [22], color moments [27] and color correlograms [28]. In our approach a normalized quantized histogram is proposed to create the feature vector of an image. Bearing in mind the performance of our model, the computational time of the retrieval process is one of the keystone design parameters. Thus, it is appropriate to reduce the color space in order to provide fast indexing even in the cases of use of big databases. This leads to incorporation of a small set of colors. In the present study 25 perceptual colors are selected from the RGB color space. These 25 colors divide uniformly the RGB space and are proposed in [22]. Table 1. presents the quantized RGB color palette. It is worthy to mention that additional features may be included in the proposed approach. However, our goal is to investigate the performance of the model by taking into account only the color of the interested points avoiding the potential noise from the rest pixels (background) and maintaining the small size of the feature vector.

Table 1. Color-lookup table.

Color	R	G	B	Color	R	G	B
Black	0	0	0	Plum	146	109	0
Sea green	0	182	0	Teal	146	182	170
Light green	0	255	170	Dark red	182	0	0
Olive green	36	73	0	Magenta	182	73	170
Aqua	36	146	170	Yellow green	182	182	0
Bright green	36	255	0	Flouro green	182	255	170
Blue	73	36	170	Red	219	73	0

Green	73	146	0	Rose	219	146	170
Turquoise	73	219	170	Yellow	219	255	0
Brown	109	36	0	Pink	255	36	170
Blue gray	109	109	170	Orange	255	146	0
Lime	109	219	0	White	255	255	255
Lavender	146	0	170				

After the recognition of the interested points of the image, the quantized image Q is created and the normalized quantized histogram is calculated. Given a color space containing L (25) color bins, the color histogram of image Q is expressed by:

$$H(x) = \frac{n_x}{N} \ (x = 1,2,\dots,L) \tag{21}$$

where n_x is the total number of interested points in the *xth* color bin and N the total number of interested points of the examined image.

4   Experimental results

In this paper a robust PIR is proposed based on color histogram of the interested points of product images. A brief description of the process can be summarized as follows:

1. The RGB query image is transformed to HSV color space. The Value channel of the query image is selected for the preprocessing filtering step.
2. The BEMD is applied to the Value channel and the first IMF component is extracted.
3. An adaptive local threshold method is utilized for the recognition of edge points.
4. The entropy is selected as the criterion for the definition of interested points.
5. The query image is quantized by utilizing the color lookup table.
6. The normalized color histogram of the interested points is created.
7. The similarity between color images is computed by using the color histogram of interested points. The Tanimoto coefficient [29] is used as a metric of the distance.

The proposed low level feature vector has been incorporated in the product recognition platform named "*Agora*" which is the outcome from the cooperation of R&D and Software department of *ChannelSight*. Our tests derived from a subset of our database [30] containing currently 48.071 Philips product images for the region of Great Britain. The database contains images from different product categories (e.g. ironing, cooking, headphones, etc.) and regions. The PR system can be applied to several types of products from different brands. In our PR system a text analysis tool is also integrated to validate and recognize products from retrieved images provided by the proposed PIR algorithm. However, the scope of the proposed paper is to examine the effectiveness of the PIR technique. Only the first 10 images from the retrieval process are taken into consideration for the recognition of query products and the product is considered as recognized only if an image related to the query product is included in the first ten retrieved images.

Eight categories from Philips's product catalog are selected and a set of query images for each category is created from products appeared in Amazon retailer. Table 2 presents the selected categories, the number of existing images in Philips's product catalog and the percentage of retrieved products.

Table 2. Details of the results.

Categories	Amazon Products	Retrieved Products	Percentage of correct retrieval	Images in Product Catalog
Ironing	37	34	91.89 %	327
Shavers	23	21	91.30 %	266
Cooking	11	11	100 %	63
Food Preparation	26	25	96.15 %	149
Coffee	21	21	100 %	480
Headphones	91	89	97.80 %	533
Audio	27	26	96.29 %	351
Hair Removal	23	22	95.65 %	137

Fig. 3. Results from Ironing and Face-Shavers category. The left image of each pair is from Philips's catalog.

Fig. 4. Results from Cooking and Headphones category.

From the experimental results it can be seen that the proposed method is resilient and can retrieve products that have some variations in color like the first pair in Fig. 3 and the last pair of images in Fig. 5. The color histogram, which is utilized as a feature for indexing images, is a statistical feature with low computational cost. It is also invariant to rotation and scale of image scene and to any displacement of objects in the image. Experimental results reveal the effectiveness of the proposed method in those cases. Moreover, it is worthy to mention the strength of the proposed technique in

cases like the first and the penultimate example in Fig. 5. From this figure it can be seen that the stamp "Promix" in the first example and the shadow of product in the third pair of images does not drastically affect the color histogram, since the only pixels that are taken into account for the creation of the normalized and quantized histogram are the interested points of an object and not the entire object. Fig. 6 presents the edges, the quantized images and the extracted interested points of the penultimate pair of images in Fig. 5. Fig. 7 presents examples in different cases such as rotation, viewpoint changes and illumination variations. Fig. 8 shows the normalized-quantized color histograms. The first histogram refers to the Amazon's image while the second histogram refers to the image appeared in the Philips product catalog. From the second histogram it can be seen that the quantity of black color is increased, however, the black color is still limited to low values due to the introduction of notion of interested points. The orange colored region reveals the quantity of white in the tested images.

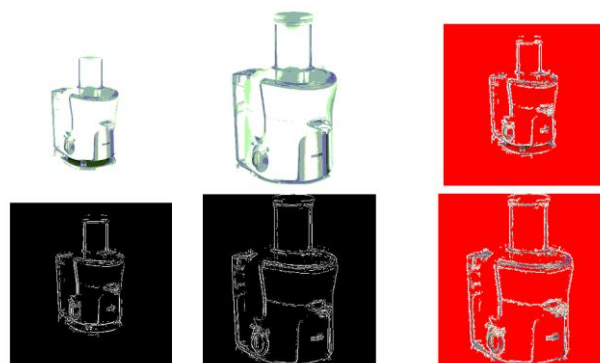**Fig. 5.** Results from Food-Preparation category.

Fig. 6. The quantized images, the edges and the interested points.

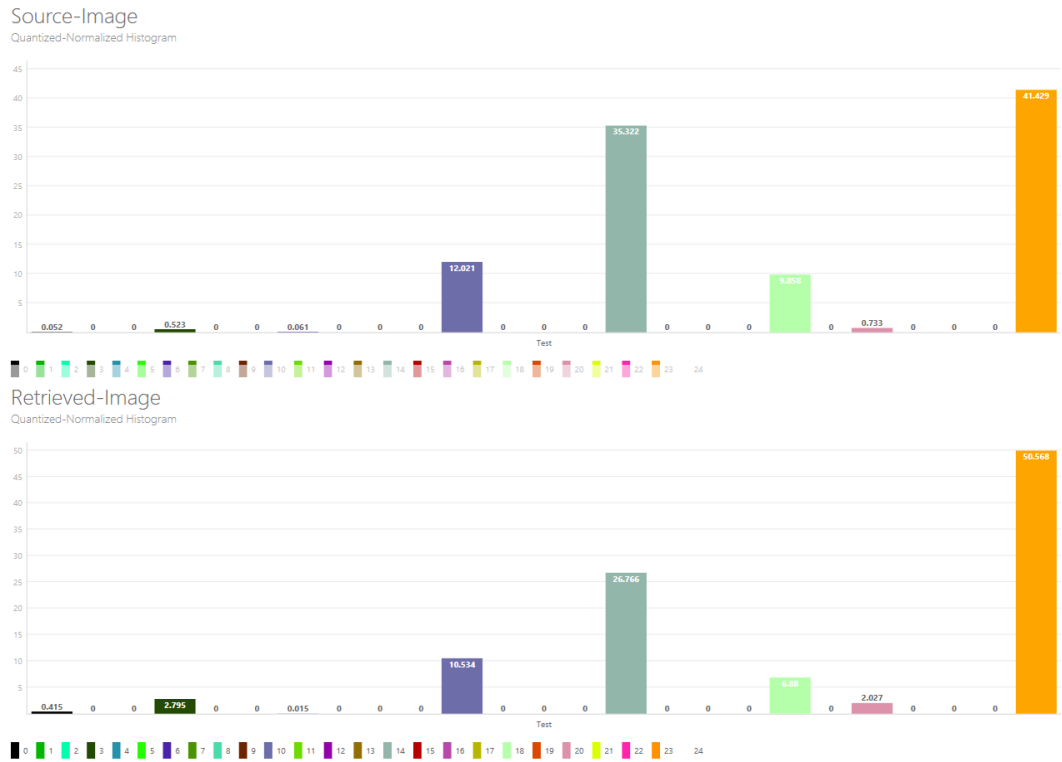**Fig. 7.** Results from Audio, Coffee and Hair removal category.

Fig. 8. Normalized and quantized color histograms for the query and retrieved image in the penultimate pair of images in Fig. 7.

Table 3. Execution time for image representation.

Size of image	64 x 64	128 x 128	256 x256	512 x 512
Execution time (s)	0.031214	0.056614	0.186249	0.678315

The total retrieval time is not constant and depends on the number of products included in the database for the specified category. It should be mentioned that the images that appeared in the product catalog are already represented as vectors and the only time load is due to the estimation of Tanimoto distances between query image and product images included in specified product category.

To validate the effectiveness of the proposed method, a comparison study between the proposed method and a number of existing CBIR in the literature is given in Fig.9. The Scalable Color Descriptor (SCD), the Edge Histogram (EH) and the Fuzzy Color and Texture Histogram (FCTH) [16] are used. SCD and EH are part of MPEG-7 specification for multimedia content description [31]. In our tests, in the case of SCD 256 coefficients are used and the number of discarded bitplanes was set to 0. The size of feature vector for SCD and EH is 256 and 80 respectively. The FCTH method provides a set of 192 features. Fig. 9 reveals the performance of each CBIR in the examined Philips product categories. From this figure it can be seen that the product recognition performance of the CBIR methods is similar except the case of Headphones where the database is larger. In this case the proposed method and the SCD provide better results. However, it should be mentioned that the performance of the proposed method is attained with the smallest size of features.
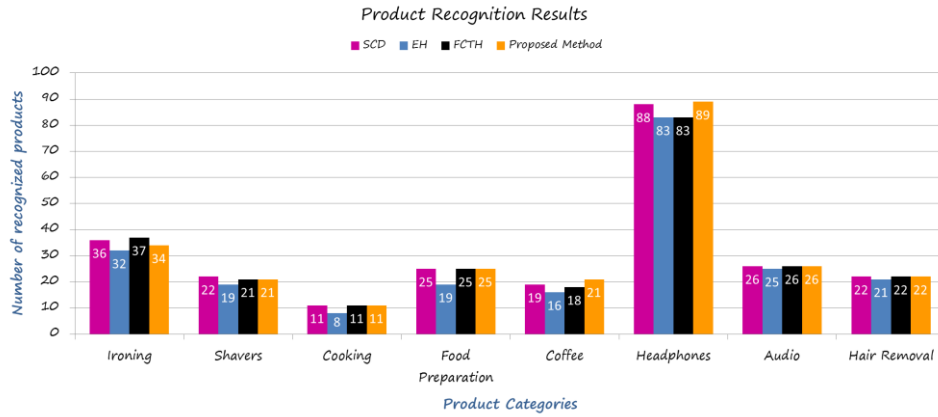

Fig. 9. Product Recognition results of the examined methods in five product categories of Philips.

5   Conclusion

In this paper a new PIR technique is proposed for the recognition of products in the e-commercial web sites. The proposed method filters the Value channel of HSV color space and extracts the first IMF through the BEMD. The first IMF provides the highest frequency content, thus it is the most informative component. An adaptive algorithm with local thresholds is applied to the first IMF for the detection of qualitative

edges. The interested points are defined following the condition of maximum entropy and the color histogram is created based on a subset of perceptual twenty five colors. The experimental results reveal the effectiveness of the proposed method. The PIR returns the associated product providing an average correct recognition over 90%. The execution time of query images with different sizes reveal that the proposed method is a real-time technique while the short feature vector is in agreement with requirements for design a fast image retrieval system. Further research involves the comparison of the proposed technique with conventional image retrieval methods in the application of product image retrieval. Moreover, future work will focus on the challenge of creating an effective and robust distributed system for product recognition, integrating also adaptive techniques that combine text analysis algorithms and CBIR methods.

References

1. Google goggles. https://support.google.com/websearch/topic/25275
2. Vuforia. https://www.vuforia.com/
3. Flow. http://www.a9.com/whatwedo/mobile-technology/flow-powered-by-amazon/
4. Jing, Y., Baluja, S.: Pagerank for product image search. In: 17th International World Wide Web Conference (IW3C2), pp. 307-316, ACM, Beijing, China, April 2008.
5. Tsai, S.S., Chen, D., Chandrasekhar, V., Takacs, G., Cheung, N.M., Vedantham, R., Grzeszczuk, R., Girod, B.: Mobile product recognition. In: International conference on Multimedia. pp. 1587-1590, ACM, Firenze, Italy, October 2010.
6. Zhang, D., Yap, K.H., Subbhuraam, S.: Mobile product recognition with efficient Bag-of-Phrase visual search. In: 6th International Symposium on Communications, Control and Signal Processing (ISCCSP), pp. 65-68, IEEE, Athens, Greece, May 2014.
7. Chai, L., Qin, Z., Zhang, H., Guo, J., Shelton, C.: Re-ranking using compression-based distance measure for content-based commercial product image retrieval. In: 19th IEEE International Conference on Image Processing (ICIP), pp. 1941-1944, IEEE, 2012.
8. He, J., Lin, T.H., Feng. J., Chang, S.F.: Mobile product search with bag of hash bits. In: 19th ACM international conference on Multimedia, pp. 839-840, ACM, Scottsdale, Arizona, USA, December 2011.
9. Wang, M, Li, G., Lu, Z., Gao, Y., Chua, T.S.: When Amazon meets Google: Product visualization by exploring multiple Web sources. In: ACM Transactions on Internet Technology (TOIT), vol. 12, no.4, 2013.
10. Eason, G., Noble, B., Sneddon, I.N.: On certain integrals of Lipschitz-Hankel type involving products of Bessel functions. In: Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol. A247, no. 935, pp. 529-551, April 1955.
11. Liu, G.H., YANG, J.Y.: Content-based image retrieval using color difference histogram. Pattern Recognition, vol. 46, no. 1, pp. 188-198, 2013.
12. Chen, K., Hennebert, J.: Content-Based Image Retrieval with LIRe and SURF on a Smartphone-Based Product Image Database. In: Pattern Recognition. Springer International Publishing, pp. 231-240, 2014.
13. Bay, H., Ess, A., Tuytelaars, T., van Gool, L.: Speeded-up robust features (SURF). In: Computer vision and image understanding, vol. 110, no. 3, pp. 346-359, 2008.

14. van Zijl, L.: Content-Based Image Retrieval with Cellular Automata. In: Cellular Automata in Image Processing and Geometry, Springer International Publishing, pp. 147-162, 2014.
15. Banda, J. M., Schuh, M. A., Wylie, T., McInerney, P., Angryk, R. A.: When too similar is bad: A practical example of the solar dynamics observatory content-based image-retrieval system. In: New Trends in Databases and Information Systems, Springer International Publishing, pp. 87-95, 2014.
16. Chatzichristofis, S.A., Boutalis, Y.S., Lux, M.: Combining color and spatial color distribution information in a fuzzy rule based compact composite descriptor. In: Communications in Computer and Information Science, Springer-Verlag, CCIS 129, J. Filipe, A. Fred, and B. Sharp (Eds.), pp. 49-60, 2011.
17. Huang, N.E., et al.: The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis. In: Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, vol. 454, no. 1971, pp. 903-995, 1998.
18. Liang, L., Ping, Z., Liu, Z.: Edge detection via window empirical mode decomposition. In: 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), IEEE, pp. 1783-1786, May 2012.
19. Guanlei, X., Xiaotong W., Xiaogang, X.: Improved bi-dimensional EMD and Hilbert spectrum for the analysis of textures. Pattern Recognition, vol. 42, no. 5, pp. 718-734, 2009.
20. Bhuiyan, S., Adhami, R., Khan, J.F.: Fast and adaptive bidimensional empirical mode decomposition using order-statistics filter based envelope estimation. EURASIP Journal on Advances in Signal Processing, 164, 2008.
21. Gatos, B., Pratikakis, I., Perantonis, S.J.: Adaptive degraded document image binarization. Pattern recognition, vol. 39, no. 3, pp. 317-327, 2006.
22. Wang, X. Y., Wu, J. F., Yang, H.Y.: Robust image retrieval based on color histogram of local feature regions. Multimedia Tools and Applications, vol. 49, no. 2, pp. 323-345, 2010.
23. Herodotou, N., Plataniotis, K.N., Venetsanopoulos, A.N.: A color segmentation scheme for object-based video coding. In: IEEE Symposium on Advances in Digital Filtering and Signal Processing, IEEE, pp. 25-29, June 1998.
24. Jain, A.: Fundamentals of Digital Image Processing. Prentice-Hall, Englewood Cliffs, NJ, 1989.
25. Sauvola J., Pietikainen, M.: Adaptive document image binarization. Pattern Recognition, vol. 33, no. 2, pp. 225–236, 2000.
26. R.J. Schilling, "Fundamentals of Robotics Analysis and Control," Prentice-Hall, Englewood Cliffs, NJ, 1990.
27. Stricker M. A., Dimai, A.: Color indexing with weak spatial constraints. In: Electronic Imaging: Science & Technology, International Society for Optics and Photonics, pp. 29-40, March, 1996.
28. Huang, J., Kumar, S.R., Mitra, M., Zhu, W.J., Zabih, R.: Image indexing using color correlograms. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 762-768, June, 1997.
29. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. 2nd ed. Wiley, 2001.
30. ChannelSight company information, http://www.channelsight.com
31. Manjunath, B.S., Ohm, J.R., Vasudevan, V.V., Yamada, A.: Color and texture descriptors. In: IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, no. 6, pp. 703–715, 2001.