7-15-2012

Smart Shopping Assistant: A Multimedia and Social Media Augmented System with Mobile Devices to Enhance Customers' Experience and Interaction

Vinh-Tiep Nguyen
University of Science & John von Neumann Institute, VNU-HCM, Ho Chi Minh city, Vietnam, tiep.nguyen@jvn.edu.vn

Trung-Nghia Le
University of Science, VNU-HCM, Ho Chi Minh city, Vietnam, 0812332@student.hcmus.edu.vn

Quoc-Minh Bui
University of Science, VNU-HCM, Ho Chi Minh city, Vietnam, 0812303@student.hcmus.edu.vn

Minh-Triet Tran
University of Science, VNU-HCM, Ho Chi Minh city, Vietnam, tmtriet@fit.hcmus.edu.vn

Anh-Duc Duong
University of Information Technology, VNU-HCM, Ho Chi Minh city, Vietnam, ducda@uit.edu.vn

SMART SHOPPING ASSISTANT: A MULTIMEDIA AND SOCIAL MEDIA AUGMENTED SYSTEM WITH MOBILE DEVICES TO ENHANCE CUSTOMERS' EXPERIENCE AND INTERACTION

Vinh-Tiep NGUYEN, University of Science & John von Neumann Institute, VNU-HCM, Ho Chi Minh city, Vietnam, nvtiep@fit.hcmus.edu.vn, tiep.nguyen@jvn.edu.vn

Trung-Nghia LE, University of Science, VNU-HCM, Ho Chi Minh city, Vietnam, 0812332@student.hcmus.edu.vn

Quoc-Minh BUI, University of Science, VNU-HCM, Ho Chi Minh city, Vietnam, 0812303@student.hcmus.edu.vn

Minh-Triet TRAN, University of Science, VNU-HCM, Ho Chi Minh city, Vietnam, tmtriet@fit.hcmus.edu.vn

Anh-Duc DUONG, University of Information Technology, VNU-HCM, Ho Chi Minh city, Vietnam, ducda@uit.edu.vn

Abstract

Multimedia, social media content, and interaction are common means to attract customers in shopping. However these features are not always fully available for customers when they go shopping in physical shopping centers. The authors propose Smart Shopping Assistant, a multimedia and social media augmented system on mobile devices to enhance users' experience and interaction in shopping. Smart Shopping turns a regular mobile device into a special prism so that a customer can enjoy multimedia, get useful social media related to a product, give feedbacks or make actions on a product during shopping. The system is specified as a flexible framework to take advantages of different visual descriptors and web information extraction modules. Experimental results show that Smart Shopping can process and provide augmented data in a realtime-manner. Smart Shopping can be used to attract more customers and to build an online social community of customers to share their interests in shopping.

Keywords: Augmented Reality, Multimedia, Social Media, Visual Query, Web Information Extraction.

1 INTRODUCTION

With the development of computing devices and network infrastructure, commerce has been changed rapidly since the late 1970s (Tkacz and Kapczynski, 2009). Electronic commerce (Tkacz and Kapczynski, 2009) then mobile-commerce (Tiwari and Buse, 2007) have great influence on not only business entities and customers (Tiwari et al., 2006) but also to the whole society in developed and developing countries.

Whether a customer uses traditional method, or e-commerce, or m-commerce, information related to a product is one of the most important factors that influence his or her decision to buy that product from one shop or another (Cialdini, 1998; Cialdini, 2009). Formal information officially published by vendors or retailers such as price, deals, or promotions to provide added value or incentives to customers, wholesalers, or other organizational customers. Social media content related a product or a vendor is formal or informal information generated by customers, such as personal comments, rating, or testimony on that product or vendor. Both formal and user-generated content on social networking sites are considered as effective business means to attract customers (Constantinideset al., 2012). Thus most online shopping systems for e-commerce and mobile applications for m-commerce provide friendly user-interface to display information related to a product in various media formats, such as text, images, video clips. Furthermore, these systems also provide users with rich interactions to collect users' content by posting reviews, rating, or even multimedia objects on products.

However, the rich information and interaction benefits in online or mobile shopping systems are not fully available for customers in traditional shopping malls. In a shopping center, only selective information on a product can be displayed next to that product, such as price, technical specifications, or promotions. TVs or computers can be placed only at several fixed locations in a shop to provide extra information on or interaction with products for limited number of customers at a time. Useful media and information, together with interactions, cannot be fully available for all customers at anytime and anywhere in shops.

These issues motivate our proposal of the Smart Shopping Assistant (SSA), a multimedia and social media augmented system with mobile devices to enhance customers' experience and interaction in shopping. Our proposed system can be considered as a bridge to deliver multimedia and social media with rich interaction from online commerce systems to users in traditional commerce environment. A user can access to all useful information, including official and user-generated content related to a product, just by looking at that product through a mobile device. The device is a smart prism connected with a server through wireless communication channel to detect a product based on its external visual appearance, e.g. textures or logos on a box or a bag, then display different information related on the product. The user is now also a member of an online community and can freely share his or her ideas/comments on products/vendors and multimedia objects related to products.

Smart Shopping Assistant consists of two main systems: a Visual Query Processing system to find the list of products with visual features similar to a given photo taken from a mobile device, and a Web Extraction and Integration System to collect and provide information from various data sources and websites on a product. Each system is design with flexible architecture so that its components can be replaced or updated to enhance accuracy and efficiency. Experimental results show that using SURF feature (Bay et al., 2008) with a single GPU-enabled server, a visual query can be processed in real-time among a collection of thousands of products. We adopt the semi-automatic web extraction approach (Hartmann et al., 2007) to generate extraction wrappers to collect products' data from different commerce related websites.

This paper is organized as follows. In section 2, we present the availability and benefits of commerce related information and interactions from the web. In section 3, we analyse the limitations of traditional commerce and propose the Smart Shopping Assistant. The two main systems in SSA, namely Visual Query Processing and Web Extraction and Integration systems, are presented in section 4 and 5. Experiments to evaluate the performance and efficiency of the proposed system are in section 6. Conclusion and future work are discussed in section 7.

2 COMMERCE RELATED INFORMATION FROM THE WEB

2.1 Availability of commerce related information from the web

Electronic commerce began to flourish rapidly from the end of 2000 and has great impacts on global economy and society (Chau et al. 2005). Besides numerous e-commerce systems at different levels of maturity and success worldwide, commerce-related websites, such as price comparison systems (e.g. *pricegrabber.com*), deals aggregators (e.g. *deals2buy.com*), or commerce related groups in social networks, have also been developed to promote purchases by providing customers with abundant amount of information on products and vendors (Barton, 2006), especially social media content (Armano, 2010).

The development of commerce and commerce-related websites are not only in developed countries but also in developing ones. Table 1 illustrates our survey on popular commerce and commercial-related websites in Vietnam (as of March 2012). We study the following features of each website: social proof; allow users post their personal ideas (comments, rating, "like") on a product; display count-down timer for promotions/deals, display number of users that purchased a product; allow online commercial transactions; and have groups in social networks (Facebook, Google+, Twitter, Zing Me – a popular social network in Vietnam).This survey is to show that even in a developing country that is not fully ready for e-commerce and m-commerce, online websites related commerce have been seriously considered and developed. The benefits of these features in marketing and commerce conversion are presented in Section 2.2.

Website	**Social proof**	**User's comments**	**Countdown timer**	**Number of purchases**	**Online purchase**	**Groups in social network[(*)]**
deal.giaothoi.com		Comment	Yes	Yes	Yes	FGT
muachung.vn		Rating	Yes	Yes	Yes	
hotdeal.vn		Comment	Yes	Yes	Yes	FT
nhommua.com	Most popular items		Yes	Yes	Yes	GZ
deal.go.vn	Most popular items	Like			Yes	
khuyenmaivang.vn			Yes	Yes	Yes	
phagia.com.vn	Most popular items		Yes	Yes	Yes	
cucre.vn	Most popular items		Yes	Yes	Yes	
runhau.com			Yes	Yes	Yes	FT
vndoan.com			Yes	Yes	Yes	
necdeal.com			Yes	Yes	Yes	F
dealtravel.vn			Yes	Yes	Yes	
muamoingay.com			Yes	Yes	Yes	FT
51deal.vn			Yes	Yes		
uudaigia.com	Most popular items		Yes	Yes	Yes	FT
hoishopping.com	Most popular items		Yes	Yes		FT
datmua.vn				Yes		FT
nhanhmua.vn			Yes	Yes	Yes	FG
windeal.vn	Most popular items	Comment	Yes	Yes		FGZ
diadiemvang.net	Most popular items		Yes	Yes	Yes	FT
everyday.vn			Yes			FT
cungmuasam.net	Having related products		Yes	Yes	Yes	FGZ
hoadeal.vn			Yes	Yes		FGT
khuyenmaisoc.vn		Comment	Yes	Yes		
giamgiakia.com			Yes	Yes		FTZ
saigonmua.vn	Most popular items	Comment	Yes	Yes	Yes	GZ
sieumua.com		Comment	Yes	Yes	Yes	FGTZ
68nhanh.com		Comment	Yes	Yes		
xdeal.vn	Most popular items	Like		Yes		FGT
5giay.vn		Comment				FGT
chodientu.vn		Comment, Like			Yes	FGT

Notes: [(*)]F: Facebook, G: Google+, T: Twitter, Z:Zing Me(Vietnamese social network)

Table 1. Popular commerce related websites in Vietnam (as of March 2012)

Commerce related information and interactions are being used successfully in both e-commerce and m-commerce to increase commerce conversion. This motivates our proposal of providing augmented reality (Furht, 2010) with multimedia and social media content on mobile devices to enhance users' experience and exciting in traditional shopping in physical stores. The availability of numerous data sources from various commerce related websites, even in developing countries, such as Vietnam, ensures the feasibility for our proposed system to automatically collect and provide useful information related to a given product for a user via a mobile device.

2.2 Influence in commerce with commerce related information from the web

We analyse important benefits of commerce related information from the web in shaping consumer behaviour with the key principles of persuasion by Robert Cialdini (Cialdini, 1998; Cialdini, 2009)

- **Social proof:** When being uncertain about something, such as the differences between two similar products, a customer often looks at others' decisions to determine because he or she believes in what other people think is correct (Cialdini, 1998; Cialdini, 2009). Thus e-commerce websites or m-commerce applications usually display "*most popular items*", "*top sellers*"... to attract and persuade users to make purchases (Amblee and Bui, 2011). In Table 1, 27/31 sites in Vietnam display the number of customers who bought a product to persuade other users to purchase and 11/31 sites show social proof data (such as "*most popular items*").
- **Authority:** Online shoppers usually use and highly regard reviews by others, especially specialists, because of their "expertise and authority" (Cialdini, 2009). Product reviews posted on blogs or social networks, or video clips in Youtube have important impacts on one's decision to purchase products. If a majority of other customers agree upon the high quality of a product, a single user may decide to purchase that product even when he or she does not really need that product.
- **Scarcity:** The announcement of the valid period of a promotion or the limited quantity of a product can boost a user's decision to buy a product because of the fear of potential loss (Cialdini, 2009). Thus commerce related websites usually display countdown timers on promotions/deals or alerts on limited quantity products. Table 1 shows that 26/31 websites in Vietnam have implemented this feature (coundown timer).
- **Liking:** Customers can be persuaded easily by other people that they like or have good relationships (Cialdini, 2009). Studies in human dynamics shows that efficiency of advertising via social networks can be optimized if good relationships between users can be exploited. Therefore most commerce related websites create online social groups in popular social networks not only to collect abundant amount of user-generated content on products but also to take for free advantages of social interactions between users sharing common interest. Table 1 shows that 19/31 websites have groups on Facebook, 11/31 on Google Plus, 15/31 on Twitter and 6/31 on Zing Me (a popular social network in Vietnam).
- **Reciprocity:** Upon knowing a good deal, people usually want to share and broadcast this news with others, especially within social groups. This can be explained by the fact that people tend to return a favour (Cialdini, 2009). Similarly, when purchasing a good product, customers usually express their personal comments or testimony on vendor's websites or social networks. This information can be used for marketing purpose as deals, promotions, and comments can be widespread easily over the network (Rosemann et al., 2011). This is another benefit of using social networks for commerce promotion.

Currently these principles are being applied widely in e-commerce and m-commerce system to boost sales. However not all of the current methods and technologies to attract and persuade customers in e-commerce/m-commerce have been exploited for traditional commerce. Some limitations of traditional commerce in physical stores are discussed in Section 3.1 as the motivation of our proposal of Smart Shopping Assistant (c.f. Section 3.2).

3 OVERVIEW OF SMART SHOPPING ASSISTANT

3.1 Limitations of traditional shopping in physical stores

With computing devices and network communications, new methods for marketing and to promote sales can be applied in e-commerce and m-commerce. However not all new features that help the success of e-commerce and m-commerce have been used to enhance users' experience in traditional shopping in physical stores. Here are several examples of limitations when a customer goes shopping in a physical store:

- Only selective information on a product is available to customers. Such information is selected subjectively by the shop owner and may not have strong influence on customers as user-generated content on a product does. Although some shops may equip screens or computers to provide more information or deals on products, these devices are located only at several fixed locations and can only serve limited users at a time.
- A customer can only see the external appearance of products, usually in boxes. That customer cannot read the table of contents of a book, watch the trailer of a DVD or a game, or watch a movie clip on user manual of a product.
- A customer cannot post comments, rating, photos or clips on product right before a shelf in a shop in a convenient way. The customer has to wait until going back home to open the web browser to post his/her testimony on products or vendors. This may prevent a customer from generating user content which is valuable in social community and has strong influence on other customers.

3.2 Proposal of Smart Shopping Assistant

The authors propose Smart Shopping Assistant (SSA) to enhance customers' experience and interactions in traditional commerce with multimedia, social media content, and other commerce related information on products. SSA aims to provide users with the following types of information:

- Commerce related information: detailed technical specifications, price at different shops/vendors, deals, sales, promotions…
- Multimedia content: photos, audio/video clips…related to a product
- Social media content: comments, rating, "like", photos, audio/video clips by users/customers...

SSA also enables users to interact with a product: posting comments, rating, "like", audio and video clips, broadcasting deals or promotions on a product…

The typical scenario of usage is as follows:

- When shopping in a physical store, a user *U* can use his/her mobile device to capture the external appearance of a product (the logo, or the front side of a box/bag). For simplicity, the term "product logo" is used in this article to represent either the logo of the front side of a box/bag containing that product. Then the user sends a "visual query image" I^*, the captured image of a product (or several products), to the Visual Query Processing system.
- Upon receiving a visual query image I^*, the Visual Query Processing system selects the list of products with logos that are similar to those in I^*. The list of products is returned to the user *U*.
- The user *U* selects a product *P*in the result list and submits the request for information related to this product to the Web Extraction and Integration System.
- The Web Extraction and Integration System executes web extraction modules corresponding to different data sources (commerce related websites or social networks) to get information related the query product *P*. All records of information are sent back to and displayed on user's mobile device.
- The user can post his/her personal comments, rating, photos, audio/video clips on a product *P*.

4 VISUAL QUERY PROCESSING SYSTEM

4.1 Visual product query

Text is the most common way to input a query into a system. However we employ the visual query system for SSA to provide a customer with comfortable means to query for products during shopping in traditional stores.

In order to detect or recognize a product in an image, we can use special patterns such as barcodes (E. Ohbuchi, 2004) or QR codes (Belussi and Hirata, 2011). However it would be more natural for a user to see or select visual information just by the regular appearance of a product rather than artificial generated pattern. Furthermore as a product logo is usually larger and easier to be recognized than a special pattern, a user can capture the product logo more easily than the special pattern. Therefore we follow the approach to detect a product based on its natural appearance.

In SSA, we do not perform the logo recognition but detection. This means that the result of a visual product query in SSA is not a single product that best match a given query image *I** but may contain multiple products with logos that are similar to those in *I**. The first reason is that there are a lot of products that have similar visual features but different technical specification. For example in Figure 1, the boxes of the two iPhones are almost the same, except for minor features ("3G" and "3GS") in the side view. When a user only captures in the visual query image *I** the external region that is identical between the two products, we cannot tell which is which. The second reason is that a user may want to find products that are similar to a given product (just like the function "*More like this*" in eBay), we should return a collection of similar products rather than the best match product.

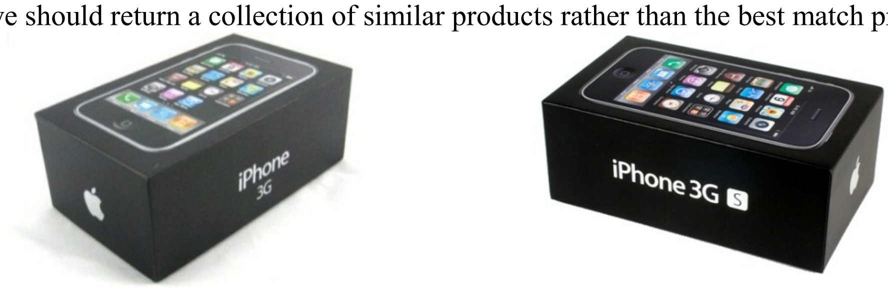


Figure 1. External visual appearance of the two products (in boxes) is almost identical

The Visual Query Processing includes two steps: lightweight filtering (c.f. Section 4.2) and product matching by visual features (c.f. Section 4.3). As there can be many products, it is necessary to filter out product logos that cannot match a query image *I** with extremely fast heuristics. Only the candidate product logos are carefully verified to evaluate the similarity with the query image *I**.

4.2 Lightweight filtering

Objective: It is time consuming to perform the detail matching between a given visual query photo with each product in the collection that may contain up to thousands of products. Therefore the lightweight filtering is used to filter out products in database that cannot match a visual query photo with low computational cost. This is the pre-processing step for the visual query process.

Approach: A product usually has main dominant colors in a logo/external appearance. In Figure 2, four products with various main distribution of colors. Based on this difference, a customer can distinguish easily a product from others. We decide to apply this approach to the Lightweight Filtering module in SSA. The color distribution of an image is considered its lightweight feature. By using this lightweight feature, in our current implementation of SSA, the lightweight filtering module can evaluate the dissimilarity between two images.

Figure 2. Four products with main dominant colors in external appearance.

Although RGB color model is the most common way to encode colors, it is not appropriate to represent the photosensitivity of human. Thus we use the HSV model because of its similarity to human color perception. A product can be in different brightness conditions, it is necessary to reduce the effect of brightness in comparing two images. Thus we only use the Hue component to create lightweight visual feature of an image.

Let Max_H be the maximum value of *Hue*. In practice, $\mathrm{Max}_H = 360$ (degree). Let $n_H > 0$ be the total number of bins for Hue channel. We calculate the n_H-bin histogram of Hue channel for each image as follows:

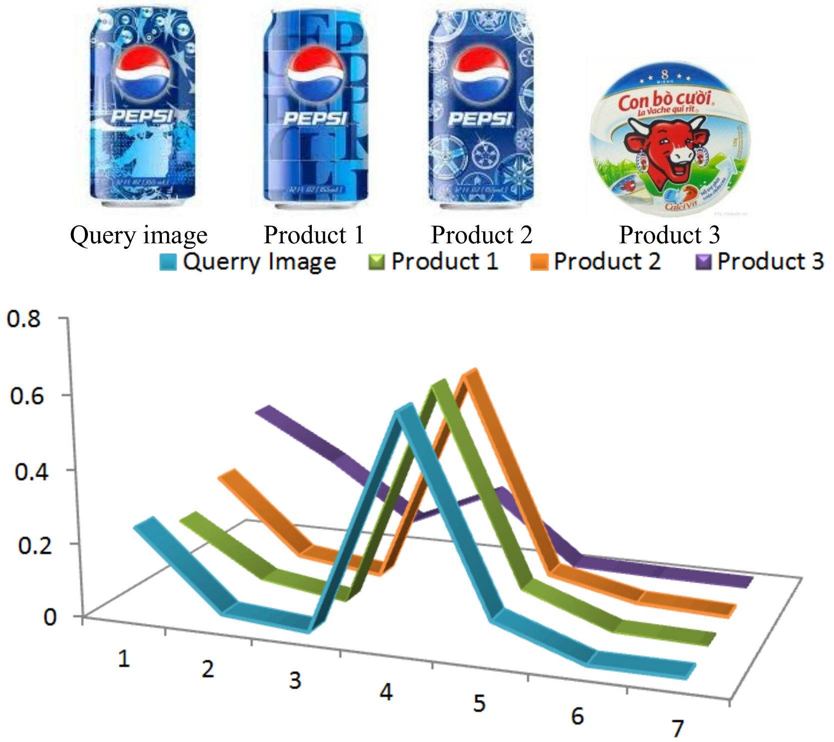
$$H_k(I) = n_k(I) / n_I \text{ for } 0 \le k < n_H$$

where $n_k(I)$ is the number of pixels in I with Hue value in $[k\ \mathrm{Max}_H / n_H, (k+1)\ \mathrm{Max}_H / n_H)$ and n_I is the total number of pixels in I. The n_H-bin histogram of Hue channel of an image is its lightweight feature.

To evaluate the difference between two images, we use a dissimilarity measure between features of a query image and a logo image. There are many dissimilarity measures but they can be classified into two main categories: bin-to-bin distance and cross-bin distance (Pele and Werman, 2010). Sum of Squared Difference (SSD) and Sum of Absolute Difference (SAD) are very widely used in many applications because it is very simple and easy to compute. They are both bin-to-bin distance so it is very sensitive to quantization of histogram. The robustness and distinctiveness of a measure depends on the way to choose number of bins in a histogram. Moreover, bin-to-bin approach only uses the relation of two corresponding bins but do not take into account the nearby ones so it can reduce the accuracy when there are some noises such as: light and contrast change in the image. In order to achieve both the robustness and distinctiveness, we use a cross-bin distance such as Quadratic-Chi histogram distance (Pele and Werman, 2010). The Quadratic-Chi histogram distance between a query image I^* and a logo I_k is determined as follows:

$$QC_{n_H}^{A}(I^*, I_k) = \sqrt{\sum_{i,j} \frac{\left(H_i(I^*) - H_i(I_k)\right)\left(H_j(I^*) - H_j(I_k)\right) A_{ij}}{\left(\left(\sum_c (H_c(I^*) + H_c(I_k))\right) A_{ci}\right)^{n_H} \left(\left(\sum_c (H_c(I^*) + H_c(I_k))\right) A_{cj}\right)^{n_H}}}$$

where $A_{i,j}$ is the similarity between bin i and j. $A_{i,j}$ get the lower value when the distance between bin i and j is farther. Distance of bin i and j is the round distance. For example, in a histogram of 255 bins, the distance between bin 1 and bin 2 is equal to the distance between bin 1 and bin 255. We use that distance because in HSV color space, especially in Hue channel, its value is an angle so 360 degree is the same with 0 degree. So to make result more accuracy, we use round distance as a measure to compute similarity matrix. We calculate the distance between a given query image I^* with each product image I_k. A product with distance less than a threshold θ_H is considered as a candidate product and we select no more than n_K candidate products with smallest dissimilarity. Figure 3 shows an example with a given query image and three products. Although the images of product 1 and product 2 are not exactly the same as the query image I^*, these two products can be considered to have similar appearance to I^* from a user's perception. In this figure, the histogram of I^* is similar to those of product 1 and product 2. These three histograms have similar distributions with the peaks at bin 5. As the histogram of product 3 is different from that of I^*, this product cannot be considered as candidate for the next step.



Histograms of the query image and three products

Figure 3. Lightweight filtering with a query image and three products

This filter can be replaced by another new filter algorithm that is more accurate and faster because each step is independent.

4.3 Product matching by visual features

Objective: The objective of this step is to verify if each of the candidate ***product logos*** found in the lightweight filtering step can be accepted as the result of the visual query process. Because the main goal of Visual Query System is not to recognize the product in a query image I^* but to select products with similar visual appearance to I^*, we can apply approaches in template matching to solve this step.

Approach: Template matching is the problem to find a sub-template in an image. There are two main approaches: area based and feature based template matching. Area based approach uses color information of a template as global features to determine the similarity between a template and an extracted pattern from a source image. There are many similarity measures such as Sum of Squared Differences (SSD), Sum of Absolute Differences (SAD), Cross-Correlation (Lewis, 1995). These methods are simple, easy to implement, and can perform with less texture objects. However they are not robust with scale, rotation transformation, and change of view point.

Feature based approaches use local features such as edges (Shi and Tomasi, 1994; Harris and Stephens, 1998), corners (Canny, 1986), blobs (Lowe, 2004; Bay et al., 2008) and a similarity measure to find the best match between local features in a template image and a source image. Because these methods are robust with scale, rotation, and change of view point, they are suitable for matching an image in which a product can be captured in different scales, poses, and orientations. Furthermore, a template for each product is large enough and has sufficient texture for this approach.

Each candidate product logo is considered as a template T to be detected in a query image I^*. If the template T is found in I^*, the corresponding product is accepted as one result of the visual query. The process to match a template T in an image I^*includes two steps: key point extraction in I^* and key point matching between T and I^*.

Step 1 – Key point extraction

In this step, key points are extracted from the query image I^*. Each key point is a blob-like structure described by its center and the properties of its neighbor region. In SIFT (Lowe, 2004), each key point is described with the descriptor of 128-dimensional vector while in SURF (Bay et al., 2008), the descriptor of a key point is just a 64-dimensional vector.

Step 2 – Key point matching between a template *T* and a query image *I**

Let Ω_T and Ω_{I^*} be the key point collections T and I^* respectively. This step includes the following steps:

- **Step A:** We find each key point p in Ω_T correspondence to key point q in Ω_{I^*} by using the nearest neighbour search. The pair *(p, q)* is called a match and is only valid if the distance between p and q is not greater than a threshold θ_M. Let Ψ_{T,I^*} be the set of collection of key points pairs *(p, q)* matches between Ω_T and Ω_{I^*}.
- **Step B:** Find the transform M than can map most of the key points in Ω_T into Ω_{I^*}. In this system, we use RANSAC method (Fischler and Bolles, 1981). This method finds the homography transform M in mapping key points. In RANSAC, we calculate the homography transform M from a subset of Ψ_{T,I^*} with randomly selected of matches (not less than 5 matches) between two images, then count the number of outliers, i.e. matches that cannot match between Ω_T into Ω_{I^*}. This step repeats until the number of iterations reaches a threshold.
- **Step C:** Assume that we have M_0 be the best homography transform with minimum number of outliers found in RANSAC process. If the number of outliers corresponding to M_0 is less than a threshold θ_H, the template T is accepted as one result of the query image I^*. Otherwise, the template T can be considered as conditionally accepted if the number of matches $|\Psi_{T,I^*}|$ is greater than a given threshold. If the number of matches $|\Psi_{T,I^*}|$ is lower than the threshold, we cannot choose template T is a result of key point matching of visual query I^*.

Figure 4. Detect a product logo in a query image using SURF features

Although the most popular feature based method is Scale Invariant Feature Transform (SIFT) by D. Lowe (Lowe, 2004), we decide to use Speeded-Up Robust Features (SURF) (Bay et al., 2008) to recognize products because it is not only faster than SIFT but also invariant with scale, rotation, illumination and view point. In SURF, a local feature (or key point) is not a single pixel but a blob-like structure which is described by its center and the properties of its neighbour region. To further speed-up the matching process, we use the GPU implementation of SURF. The experiment to compare the performance between normal version and GPU-based version of SURF is presented in Section 6.2.

Figure 4 shows an example of template matching using SURF features with the product logo T (left) and the query image I^* (right). Each line mapping from the left to the right is a pair of corresponding SURF features. This example also emphasizes that for a give query image, multiple products can be detected. If a user wants to query only a single product, he or she can capture only that product in the query image.

5 PRODUCT INFORMATION EXTRACTION FROM THE WEB

5.1 Overview of Web Extraction and Integration System

Searching for commerce related information on the web is a regular activity of Internet users (Spink and Jansen, 2008). Although there are many results corresponding to a query, a user tends to view only few first result pages (Spink and Jansen, 2004). Thus a user may not have a complete view on a given product and it is necessary to have a system to collect, summarize, and present information on a product from different data sources, i.e. websites.

To give a user with useful information from the web, we target three groups of potential data sources:

- Social networks: each social network usually provides a collection of APIs so that developers can write applications that can communicate with that social network.
- Commerce related websites with RSS: several websites provide RSS for RSS feed readers or mashup applications (Di Lorenzo, 2009).
- Commerce related websites without RSS and APIs: there are still a lot of commerce related websites that do not provide RSS or APIs to collaborate with other websites or mashup applications.

We propose to apply the plug-in mechanism to provide SSA with flexible extensions to communicate with different data sources. There are three main components in the Web Information Extraction and Integration system of SSA:

- Social Network Plugin Manager is responsible for managing plugins to communicate with different social networks (e.g. Facebook, Twitter, Google+, Zing Me). If a new set of APIs is available, we can easily create a new plugin for that social network.
- RSS Manager processes RSS data from the list of data sources that can be updated regularly.
- Web Extraction Plugin Manager is in charge of managing plugins (web extraction modules) to extract data from different commerce related websites. Each module can extract necessary data from a given website in a pre-defined scenario. As there are still many websites that do not follow the trend of Web 2.0 to provide APIs or RSS, this approach is efficient to convert a website without RSS or APIs into a data source.

The development of the first two main components are straightforward. We focus our presentation of creating web extraction modules in Section 5.2.

5.2 Web extraction module

There are different methods to develop web extraction modules, which can be classified into three main categories: manual, automatic, and semi-automatic approaches. The manual approach is not an efficient way because a developer has to write manually new extraction module for a new or updated website. In the automatic group, there are different ways to create a smart bot to extract data from a website, such as automatic data extraction based on similarity between in web pages (Crescenzi, 2001), HTML DOM structure analysis (Álvarez et al., 2010; Novotny et al., 2009; Shaker et al., 2009). These methods are efficient when we want to automatically explore, analyse, and extract data from arbitrary websites. However the accuracy of these methods is usually lower than semi-automatic approaches (Hartmann, 2007; Pan et al., 2002).

In SSA, as we manually select websites to extract commerce related information, we follow the semi-automatic approach. The key idea of semi-automatic approach is to record user’s activities in the process to search for information on the web, then replay this sequence of activities with appropriate parameters to get web data corresponding to a new query (Hartmann, 2007).

We briefly summarize the main phases of creating and executing a web bot to extract data from a website as follows:

- **Phase 1 - Web bot training:** We develop a web bot designer system to record all user's activities on a web browser, including filling text data into a textbox, selecting an option in a combobox, checking/unchecking a checkbox, selecting a radio button, navigating to a new page... This is called web bot scenario. The scenario is parameterized with text data to be filled in a textbox, the index of an option to be selected, the index of the hyperlink to be invoked… These parameters are used to customize the execution of a scenario with new contexts (or queries).
- **Phase 2 - Web bot execution:** The web bot scenario is replayed automatically with appropriate parameters. Finally the web bot module extracts required fields (text, image) in the last page of the navigation sequence using HTML DOM paths.

This approach is a natural way to capture and mimic a user's activities in searching for information related to a query in a website. Using this method, we can easily create web bots corresponding to different queries in various websites with high accuracy and in an efficient way.

6 EXPERIMENTS

In this section we present experiments to evaluate different properties of Smart Shopping Assistant system. First we conduct experiments to evaluate the performance (speed) of matching a product logo *T* with a query image *I** with two implementations of SURF: the normal version and the GPU version (c.f. Section 6.2). We then perform experiments to evaluate the efficiency of the pre-processing step (Lightweight Filtering) on the overall speed of the Visual Query Processing system (c.f. Section 6.3).Experiments on the performance of the web bot modules to extract information from different commerce related websites that do not support RSS and APIs are presented in Section 6.4. Finally we present several scenarios of usage to demonstrate the usability of the proposed system SSA (c.f. Section 0).

6.1 Datasets for experiments

For experiment in this paper, we prepare 5 datasets of product logos from 5 popular commercial corporations in Vietnam, namely Phong Vu computer, Vinabook, Hoan Long computer, Mai Nguyen, and Golmart. These corporations have both traditional shopping centers and commercial websites. Table 2 shows the product types and number of products (that we collect for experiments) of each corporation. We choose these datasets because we want to evaluate SSA with different types of products and different number of products in a dataset. Furthermore we can get commerce related information on a product right the official website of a corporation to implement the prototype of SSA.

Corporation	Product type	Number of products (**)
Vbook http://vbook.vn	Books, magazines	51
Mai Nguyen http://www.mainguyen.vn/	Cell phones	283
Phong Vu computer http://www.vitinhphongvu.com/	Computers and peripheral devices	439
Hoan Long computer http://hoanlong.com.vn/	Computers and peripheral devices	800
Golmart http://www.golmart.vn	Supermarket	1609

Table 2. Testing datasets from commercial corparations in Vietnam

Notes: (**) This is only the number of products that we collect to conduct experiments in this paper. This is not the total number of products of the shop.

6.2 Image matching using SURF features

Although matching images with SURF features is much faster than with SIFT features (Bay, 2008), it is necessary to further speed up this process to improve the overall performance of the Visual Query Processing system, especially when the system is used to serve multiple users.

Using the parallel processing capability of GPU, we can improve significantly the speed of image matching process. We use two implementations for image matching using SURF features: a normal implementation (using CPU only) and a GPU-enabled one (using both CPU and GPU). Experiments are performed in a system using Core 2 Duo P8400 2.26 GHz (with 2GB RAM) and a graphic card GeForce GTX 460 (1GB memory). This system is used for all experiments in this paper.

For each dataset, we randomly select 100 pairs of product logos and perform the image matching. Then we calculate the average time for each match. The result of this experiment is shown in Figure 5. On average it takes 30-40ms to match images using the CPU-only version and 2-3ms using the GPU-enabled version. This result shows that the GPU-enabled version outperforms CPU-only version and the image matching process with GPU-enable implementation is acceptable for a real-time image matching. In case we want to further speed-up the process for more users, we can take advantages of the parallel processing with multiple GPUs or even cloud computing architecture (Mell and Grance, 2011).

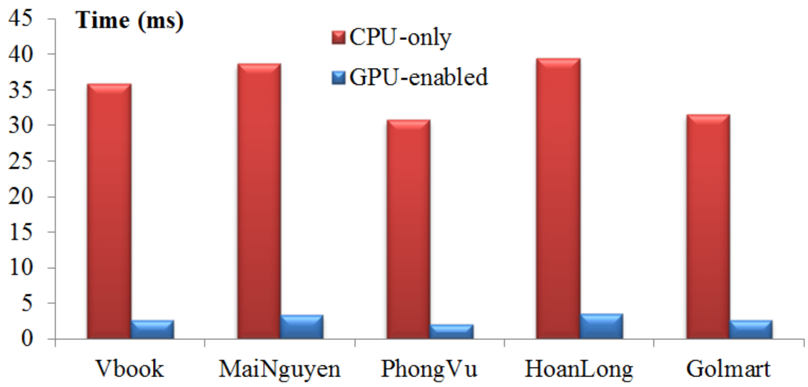


Figure 5. Average time for image matching with SURF features

6.3 Efficiency of Lightweight Filtering

In SSA, we want to choose a collection of similar matches with a give query image *I** so that Visual Query Processing system return not only a single product that best matches with a given query image but a list of product with similar logos. This experiment is set to illustrate the efficiency of the Lightweight Filtering in processing a visual query.

For each dataset, we use 50 visual queries with various input images. For each visual query, we conduct the visual query process in two scenarios: without Lightweight Filtering and with Lightweight Filtering. The results of this experiment are show in Figure 6.

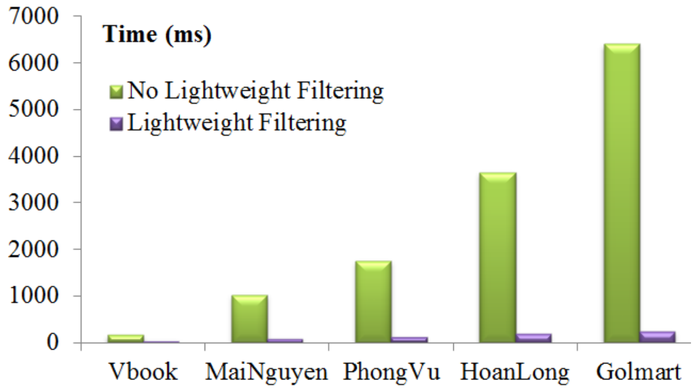


Figure 6.Comparison between the performance (in milliseconds) of processing a visual query with and without Lightweight Filtering

Scenario 1: Each query image I^* is matched with each product logo in a dataset. The result shows that the total time for processing a query image increases with the number of product logos.

Scenario 2: We only choose the top n_k candidates product logos in a dataset for matching with SURF features. In this experiment, we set $n_k = 10$. The time to process a visual query in this scenario do not change too much when the total number of product logos in dataset. The reason is image matching is only executed with no more than n_k candidate product logos for each query.

On average, it takes about 45 – 60 milliseconds for all steps of visual query process (for $n_k = 10$). The average elapsed time is slightly higher than the total time for matching a visual query image I^* with $n_k = 10$ candidate logos because of the extra time to perform the Lightweight Filtering.

6.4 Web information extraction

This experiment is independent from the previous two experiments. Our objective in conducting this experiment is to evaluate the applicability of Smart Shopping Assistant in extracting commerce related information from websites, even with sites that do not provide RSS and APIs, with the current network infrastructure of Vietnam.

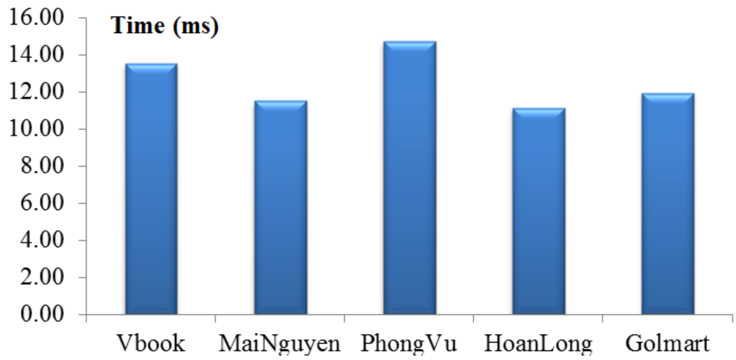


Figure 7.Average time to search and extract web information

The extracting task is done by performing the following steps by a web bot that was trained to explore a give website:

- Navigate to the homepage of a website
- Input the product's name into the search textbox
- Invoke the search function of the query form in the homepage
- Wait for response streams from the server (of the corporation)
- Analyse and extract data (text and images) in the result page.

Figure 7 shows the average time (in milliseconds) to execute a web bot to replay and extract data from a website. From this figure, it takes less than 14ms to extract information related a product from a website. Thus Smart Shopping Assistant is expected to provide information on a product from multiple data sources with less than 1 second.

6.5 Scenarios of usage

In Figure 8, we present several different scenarios of usage to illustrate the functions of our current implementation of Smart Shopping Assistant. A mobile device becomes a special prism through which a user can see the product of interest (as an image captured by the camera of that mobile device) augmented by commerce-related information (a), multimedia and social media contents (b) and (c). Furthermore a user can vote (d) or give comments (e) to a product just by tapping on the options on the screen.

a. Sale b. Photo gallery c. Video clip d. "Like" e. Comment

Figure 8. Several scenarios of usage with Smart Shopping Assistant

7 CONCLUSION

In this paper we analyse the necessity of providing extra multimedia and social media contents of products to customers when shopping in physical shopping centers. As mobile devices are becoming more popular in society, we propose the Smart Shopping Assistant that uses mobile devices as special prisms for users to capture visual features of products, to display related information about products. The proposed system also provides users with comfortable user interface to interact with products. A user can post comments, rating, photos, or video clips on a product whenever he/she wants at any place in the shopping centers. In this way, a user can freely express his/her own attitude or feeling toward a product right at the moment of excitement. In general, Smart Shopping Assistant can be considered to bridge the gaps between e-commerce/m-commerce and traditional commerce.

With the flexible architecture, Smart Shopping Assistant can be upgraded with new algorithms for object detection or recognition based on visual features. With the plug-in mechanism, new modules to extract information from various data sources in the web can be implemented and added to SSA. With the Social Network Plugin Manager, SSA can easily collaborate with various social networks to extend social reach of online groups via mobile devices.

This proposed system is our initial step to contribute to the development of a social community in commerce. By collecting users' data and interactions, we can further study useful personal interests and behaviours in shopping, and then provide recommendations for users in shopping. Furthermore, when integrating with localization feature using Global Position System (GPS), Cell ID, or Near Field Communication (NFC), SSA will become smarter with context-aware capability to guide users in shopping centers.

References

M. Álvarez, A. Pan, J. Raposo, F. Bellas, and F. Cacheda (2010).Finding and Extracting Data Records from Web Pages. *Journal of Signal Processing Systems*, vol. 59, pp. 123-137.

D. Armano (2010), Six Social Media Trends for 2011. *Havard Business Review*, available online: http://blogs.hbr.org/cs/2010/12/six_social_media_trends_for_20_1.html

N. Amblee, T. Bui (2011). Harnessing the Influence of Social Proof in Online Shopping: The Effect of Electronic Word of Mouth on Sales of Digital Microproducts. *Journal International Journal of Electronic Commerce*, Vol. 16, No. 2, pp. 91-114.

B. Barton (2006). Ratings, Reviews & ROI: How Leading Retailers Use Customer Word of Mouth in Marketing and Merchandising. *Journal of Interactive Advertising*, Vol 7 No 1, pp. 47-50.

L. F. F. Belussi and N. S. T. Hirata (2011).Fast QR Code Detection in Arbitrarily Acquired Images.In Conference on Graphics, *Patterns and Images* (SIBGRAPI). pp. 281 - 288.

H. Bay, A. Ess, T. Tuytelaars, L. V. Gool (2008). SURF: Speeded Up Robust Features. *Computer Vision and Image Understanding* (CVIU), pp. 346-359.

R. B. Cialdini (1998). Influence: The Psychology of Persuasion, Collins.

E. Constantinides, C. Lorenzo-Romero, A. M.d.C. Alarcón-del (2012). Social Networking Sites as Business Tools: A Study of User Behavior, *Advances in Business Process Management*, Springer-Verlag.

R. B. Cialdini (2009). *Influence: Science and practice* (5th edition). Boston: Allyn& Bacon. ISBN 978-0205609994.

J. Canny (1986).A Computational Approach to Edge Detection. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*. pp. 679-698.

C. E. H. Chau, H. M. Khoo, D. W. Straub, S. Kadiyala and D. Kuechler (2005). The Evolution of E-Commerce Research: A Stakeholder Perspective. *Journal of Electronic Commerce Research*, Vol. 6, No. 4: pp. 262-280.

V. Crescenzi, G. Mecca, and P. Merialdo (2001). RoadRunner: Towards Automatic Data Extraction from Large Web Sites. Proceedings of *the 27th International Conference on Very Large Data Bases*, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., pp. 109-118.

B. Furht (2011). *Handbook of Augmented Reality*. Springer, p.3-46. ISBN 978-1-4614-0063-9

M. A. Fischler, R. C. Bolles (1981). Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Comm. of the ACM*, Vol 24, pp 381-395.

B. Hartmann, L. Wu, K. Collins, S.R. Klemmer (2007). Programming by a sample: rapidly creating web applications with d.mix. Proceedings of the *20th annual ACM symposium on User interface software and technology - UIST '07*, New York, New York, USA: ACM Press, 2007, p. 241

C. Harris, M. Stephens (1998). A combined corner and edge detector. In Proceeding *Alvey Vision Conference* .pp. 147-151.

B. Hartmann, L. Wu, K. Collins, and S.R. Klemmer (2007). Programming by a sample: rapidly creating web applications with d.mix. Proceedings of the *20th annual ACM symposium on User interface software and technology - UIST '07*, New York, New York, USA: ACM Press, p. 241.

D. G. Lowe (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* (IJCV). pp. 91-110.

J. P. Lewis (1995). Fast normalized cross-correlation. In *Vision Interface, Canadian Image Processing and Pattern Recognition Society*. pp. 120 - 123.

G. Di Lorenzo, H. Hacid, H.Y. Paik, and B. Benatallah (2009). Data integration in mashups. *ACM SIGMOD Record*, vol. 38, Jun. 2009, p. 59.

P. Mell, T. Grance (2011). *The NIST Definition of Cloud Computing*. National Institute of Science and Technology

E. Ohbuchi, H. Hanaizumi, L. A. Hock (2004). Barcode readers using the camera device in mobile phones.In *International Conference on Cyberworlds*. pp. 260 – 265.

O. Pele, M. Werman (2010). The quadratic-chi histogram distance family. In Proceedings of *European conference on Computer vision* (ECCV). pp. 749-762.

A. Pan, J. Raposo, M. Álvarez, J. Hidalgo, and Á. Viña (2002). Semi-Automatic Wrapper Generation for Commercial Web Sources. *IFIP Conference Proceedings*, vol. 21, pp. 265 - 283.

M. Rosemann, M. Andersson, M. Lind (2011).Digital Complemenatary Assets. Paper presented at the *International Conference on Information Sys*tems 2011. Shanghai, P.R.C.

A. Spink, B.J. Jansen (2008). Trends in Searching for Commerce Related Information on Web Search Engines, *Journal of Electronic Commerce Research*, VOL 9, NO 2, 2008,pp. 154-161

A. Spink, B. J. Jansen (2004). *Web Search: Public Searching of the Web*. Springer.

J. Shi and C. Tomasi (1994). Good Features to Track. In *IEEE Conference on Computer Vision and Pattern Recognition*. pp. 593 - 600.

R. Szeliski (2010). *Computer Vision: Algorithms and Applications*. Springer.

E. Tkacz, A. Kapczynski (2009). *Internet – Technical Development and Applications*. Springer, p. 255

R. Tiwari, S. Buse and C. Herstatt (2006). From electronic to mobile commerce opportunities through technology convergence for business services.*Asia Pacific Tech Monitor* , pp. 38-45.

R. Tiwari, S. Buse. (2007). *The Mobile Commerce Prospects: A Strategic Analysis of Opportunities in the Banking Sector*. Hamburg University Press, pp 33.

P.A. Viola and M.J. Jones (2001).Rapid object detection using a boosted cascade of simple features. In *Conference on Computer Vision and Pattern Recognition* (CVPR).pp 511-518.