Mathematical Problems of Computer Science 39, 101–105, 2013.

Searching Images by Using Color Structure

Rafayel V. Barseghyan

Institute for Informatics and Automation Problems of NAS RA
e-mail: barseghyan@gmail.com

Abstract

In this paper a color-structured image search method is proposed which allows to classify large image collections by color and quickly find the interested images. According to the proposed method a software system is developed that demonstrates the efficiency and accurancy of the proposed method.

Keywords: image search, color structure, visual similarity.

1. Introduction

In recent years image search engines, such as Google Images [1], Yandex Images [2], Bing Images [3],TinEye [4], PicSearch [5] and many others, are being rapidly developed. These search engines allow to find the desired images by using various criteria such as creation dates, orientation, size, color and visual similiarity (duplicates).
Methods for content based image searching can be divided into two major types: visual re-ranking [6], [7] and prediction-based refinement [8]. IntentSearch [8] provides a mechanism to allow the users to select a few images of interest, and automatically mine the user's intent to reorder image search results. In this paper, a simple method for finding images based on explicit intent is presented. The developed software allows to find the desired images by using the percentage of the desired colors.
In [9] and [10] you can see a system which is similiar to the suggested method. [9] was developed and supported by Idee Labs [11] Canadian-based company which owned TinEye [4]. Multicolor Search Lab [9] makes color-based image search within more then 10 million images from Flickr [12] which has a license of Creative Commons. [9] has a control of desired colors concentration in images. The users can select up to 5 colors and control color concentration balance. As a consequence of increase in concentration percent of one color the other colors concentration percent decreases.
In [9] (see Figure 1) base colors palette used contains 256 colors. In [13] used palette contains 28 colors, Google's (see Figure 2), Bing's and Picsearch's colors palette contains 12 color and Yandex's (see Figure 3) palette contains 9 colors.

2. Method

In the first step of our algorithm we need to correspond each pixel color of every image to our color from our base colors palette. As base colors the Google's color palette is used. For

this we need to get RGB values of each pixel in our color palette. After that we need to find the closest value of color in RGB system for each pixel. As a distance metrice mean square error is used. The mathematical formula is presented below:

$$\begin{aligned} D &= (pixel_{pal}.rgb.r - pixel_{img}.rgb.r)^2 + (pixel_{pal}.rgb.g - pixel_{img}.rgb.g)^2 + \\ &+ (pixel_{pal}.rgb.b - pixel_{img}.rgb.b)^2 \\ & D- > min \end{aligned} \quad (1)$$

where $pixel_{pal}.rgb$ - is a rgb value of palette color and $pixel_{pal}.rgb.r$, $pixel_{pal}.rgb.g$ and $pixel_{pal}.rgb.b$ are, respectively, the red, green and blue values of rgb.

3. Realization

This section is divided into four logical parts: Getting the images from Flickr, indexing (analyzing) the images, storing the indexed results and searching within the stored data.

- The first step incorporates mining Flickr for images. Fortunately, Flickr provides a way of doing this: it allows the software of outside users to interact with their API [14]. We have chosen all images which are tagged as "Armenia" and licensed as "Creative Commons". Flickr search returned 16250 images meeting these criteria. For doing this with Flickr API we need to signup as a Flickr developer in order to retrieve the "API_KEY". One of the limitations of the Flickr API is that one can only retrieve a maximum of 500 images per page for a specified search term.
  It is interesting fact that the downloaded images contain only 39 images in GIF format and 43 images in PNG format, the others are coded by JPEG format.

- After downloading the required images we start to analyse them. For doing that we have written an indexer. The indexer was written by using ImageMagick [15]. ImageMagick is a widely known software library to create, edit, compose, or convert bitmap images. It can read and write images in a variety of formats (over 100) (the full list of the supported formats one can see in [16]).
  We calculate the percents of palette's colors in given image. We take only that colors which have more than one percent existance in the given image.

- The most critical part of our system is storing information of the analysed images. We can use classic SQL datastore such as SQLite, MySQL or PostgreSQL, but for having more speed when doing search we choose new types of datastore Redis [17]. Redis is a schemeless, fast and reliable database engine. Redis stands out from many other NoSQL databases with its powerfull datatypes. For our system the best choice of datatype is the "Sorted Sets".
  We store the analysed data as

  $$color\_id \qquad percent \qquad img\_id$$

  where img_id -is the identifaction number of image, $color\_id$ - is the identifaction number of color from palette and $percent$ is the percent of $color\_id$ in the given image. If the given color doesn't exist in the given image we store 0.

- For searching images with the required concentration of colors we write queries to our database. For getting images which contain the required percents of the desired color we need to operate on scores. We can do this by using the "zrangescore" operator. The operator "zrangescore" returns a set.

$$zrangebyscore \quad color\_id \quad percent \quad < \quad input\_percent$$

After getting the required sets for all desired colors we do sets intersection. To exclude colors we need to do additional query like this

$$zrangebyscore \quad color\_id \quad percent \quad -inf \quad 0$$

4. Conclusion

The developed software system can help the owners of large image archives to easily find the desired images. To get more relevant images the system supports multicolor search and allows the user to include/exclude colors and control percentage color of the desired image. The developed software system works fast, for example, the search process in our collection with 16250 images takes only 0,07 seconds.

References

[1] http://images.google.com

[2] http://images.yandex.ru

[3] http://bing.com/images

[4] http://tineye.com

[5] http://picsearch.com

[6] W. H. Hsu, L. S. Kennedy, and S.-F. Chang, "Reranking methods for visual search," *IEEE MultiMedia*, vol. 14, no. 3, pp.14-22, 2007.

[7] Y. Jing and S. Baluja, "Pagerank for product image search. In WWW," WWW 2008, Beijing, China, pp. 307 - 316, 2008.

[8] J. Cui, F. Wen, and X. Tang, "IntentSearch: Interactive on-line image search re-ranking," *Proc. 16th ACM Int'l Conf. Multimedia*, pp. 997-998, 2008.

[9] http://labs.tineye.com/multicolr

[10] http://labs.tineye.com/multicolour

[11] http://ideeinc.com

[12] http://flickr.com

[13] http://dribbble.com

[14] http://flickr.com/services/api/

[15] http://imagemagick.org

[16] http://imagemagick.org/script/formats.php

[17] http://redis.io

Figure 1: Tineye Multicolr Search Lab interface.

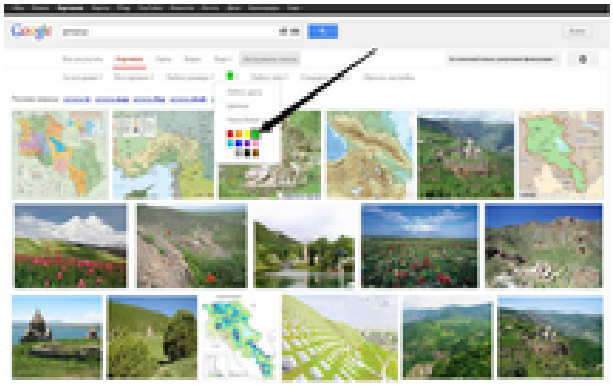

Figure 2: Google's image search interface.

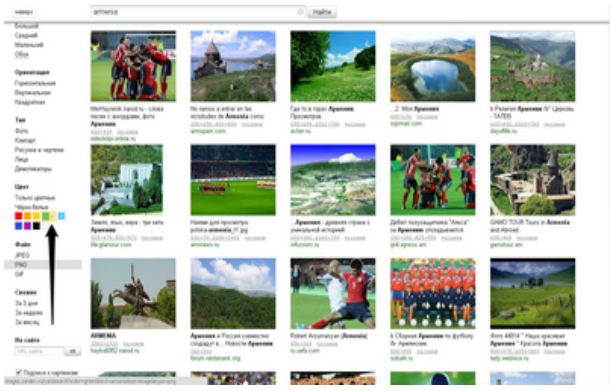

Figure 3: Yandex's image search interface.

Submitted 16.01.2013, accepted 25.02.2013.

Պատկերների որոնում ըստ գունային հատկանիշների

Ռ. Բարսեղյան

Ամփոփում

Աշխատանքում առաջարկված է մեթոդ, որը թույլ է տալիս որոնել պատկերներ՝ օգտագործելով դրանց գունային հատկանիշները: Մշակված մեթոդը կօգնի պատկերների մեծ հավաքածու ունեցողներին հեշտությամբ գտնել պահանջվող պատկերները: Մշակվել է ծրագրային համակարգ, որը՝ որպես փնտրման պարամետր, օգտագործում է ցանկալի գույնի (գույների) առկայությունը/ բացակայությունը: Համակարգը թույլ է տալիս կառավարել պահանջվող գույնի (գույների) առկայության միջակայքը: Համակարգը փորձարկվել է օգտատերերի կողմից Flickr էլեկտրոնային ռեսուրս վերբեռնված “Armenia” - պիտակով նշագրված և “Creative Commons” լիցենզիա ունեցող նկարների բազմության վրա:

Поиск изображения по цветовым характеристикам

Р. Барсегян

Аннотация

В работе предложен метод для поиска изображений по цветовым характеристикам. Разработанный метод призван помочь владельцам больших архивов изображений быстро и визуально просто найти нужные изображения. В соответствии с предложенным методом было разработано программное обеспечение, которое позволяет искать нужные изображения как по вхождению определенного цвета (определенных цветов), так и исключению определенного цвета (определенных цветов), а также позволяет задать процент вхождения цвета (цветов). Разработанное программное обеспечение было тестировано на базе изображений, которые были загружены пользователями в интернет сервис Flickr и были помечены меткой "Armenia" и имели лицензию "Creative Commons".