Color descriptor for content-based drawing retrieval

Christophe Rigaud, Jean-Christophe Burie, Jean-Marc Ogier, Dimosthenis Karatzas

HAL Id: hal-01032447

https://hal.archives-ouvertes.fr/hal-01032447

Submitted on 23 Jul 2014

Color descriptor for content-based drawing retrieval

Christophe Rigaud, Jean-Christophe Burie, Jean-Marc Ogier
Laboratory L3i
Université de La Rochelle
Avenue Michel Crépeau 17042 La Rochelle, France
{christophe.rigaud, jean-marc.ogier, jean-christophe.burie}@univ-lr.fr

Dimosthenis Karatzas
Computer Vision Center
Universitat Autònoma de Barcelona
E-08193 Bellaterra (Barcelona), Spain
dimos@cvc.uab.es

Abstract—**Human detection in computer vision field is an active field of research. Extending this to human-like drawings such as the main characters in comic book stories is not trivial. Comics analysis is a very recent field of research at the intersection of graphics, texts, objects and people recognition. The detection of the main comic characters is an essential step towards a fully automatic comic book understanding. This paper presents a color-based approach for comics character retrieval using content-based drawing retrieval and color palette.**

Keywords—color descriptor, non-rigid object, drawing retrieval, color palette, comics analysis

I. Introduction

Human detection in computer vision field has been largely studied during the past decades, mainly based on grayscale image and gradients. Color information is rarely relevant for human detection because of clothing and skin color difference. In videos, moving regions are often used as region of interest.

Although, comics are often reproduction of human life situations, they are a domain where we can not directly human detection based methods. The main difference is that comics characters (e.g. hero, second actor) are hand drawn, they are much more variant in terms of deformation, shape and appearance than real life humans (see fig. 1). Comics are composed by a sequence of static snapshots telling a story but we can imagine that a comics strip is like a video at a very low frame rate and time step. In colored comics, the color information gives the identity of characters and plays a main role for character spotting. Comics characters are often described using colors while human are described using head feature (e.g. brown hair, blue eyes) because it is the only part that does not change in every day life. This is a big difference compared to natural images, since comics are designed in a way that the information can be quickly found.

Fig. 1. Examples of comics character postures. This example shows deformation, pose, rotation, translation and occlusion variations. Image credits: Prunelle, la fille du cyclope, Vicky Portail-Kernel and Cédric Kernel, Ankama, 2010.

Recent studies have been published for partial Manga copy detection [1], mainly based on shape information because of absence of any color information. This work has been extended to Manga copyright infringement protection [2] using a face detector [3] as seeds and concentric HOG [4] descriptor. This work shows good results for Manga part retrieval which is a subset of the whole comics world. In our work we focus on a complementary subset, corresponding to colored comics analysis.

Colored comics may be compared to cartoon images sequence, for which a first work based on HOG, SVM and color attributes has been published in 2012 [5]. Another work presents the other way around from a movie to comics representation [6], using image cartoonification and script analysis. Recently, graph theory has been used to find redundant color structure in order to automatically localize the most frequent color group apparition and label them as main characters [7]. A thesis [8] (section 1.1 and 1.2) gives a good overview about hand drawing recognition in which it is mentioned that "story board scene understanding still remains an open problem, few results are available in the literature about stroke images". Another work uses HOG descriptor with redundant information classification to also find the most frequent elements [9]. Both graph and descriptor based methods need a preprocessing step to remove unwanted redundant elements such as text letters and speech balloons. This could be done by using text [10] and balloon [11] detectors.

Next section will present the proposed method, followed by experiments, discussion and conclusion.

II. Proposed method

In this section we detail how to localize comics characters in all pages of a given album. We aim at detecting the apparitions of a comics character (non-rigid object) with a minimum of user interaction. This is a content based image retrieval problem where the result highly depends on the quality of the given example, that should be noise free (no background, target only). A close up view of comics drawings shows that the most invariant information between texture, shape and color is the color information. We believe that this feature is the only information which is invariant to scale, object deformation, translation and rotation transformation which are very frequent in comics. The typical drawback of using color features is that they are sensitive to illumination and contrast variations. This is an important issue for natural images but not for hand drawings if they are drawn and digitized under the same conditions.

For all those reasons we based our method on color information only. Given a comic page, we first ask the user

about the position of one object example to detect and then perform and exhaustive search in all the pages of the same album. Note, the color selection has to be carefully done to avoid background details.

The proposed method can be summarized as follows (for a given comics album) :

- get the object colors from input query
- compute and apply a reduced color palette to all pages
- compute the query descriptor
- retrieve similar objects

The retrieval (search) step can be extended to other albums from the same collection (similar comics characters).

A. Input query

A minimal user interaction is necessary to tell the system what to retrieve. We only ask the user to give an information about the object's colors by selecting few pixels from the query object. We propose to use a pointing device (e.g. mouse, finger on touchsreen) to click, drag and release over the object colors and make the selection of pixels (see fig. 2).
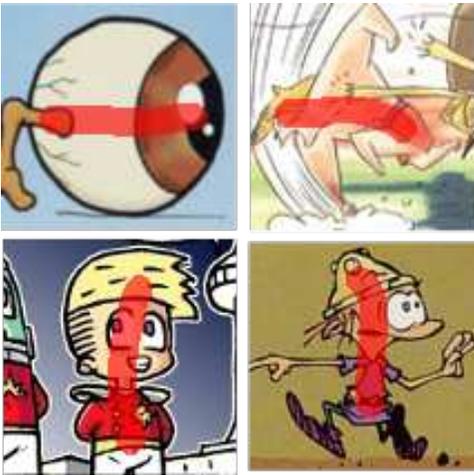

Fig. 2. Examples of comics character color selection using a pointing device (click, drag and release). Selected pixels are highlighted in red, they constitute the query example given by the user.

B. Color reduction

Once the user has defined object colors, we reduce the number of color of all the images of the album (including the query) according to a color palette P_c in order to describe and retrieve the object with few colors.

C. Query descriptor

The query descriptor we use is a non ordered color vector that contains the information of presence of certain colors in the query ($D \subset P_c$). It is not composed by spatial or quantitative information (e.g. region positions, number of pixels) because is it not scale and deformation invariant.

The color subset D is composed by N colors which have the highest number of pixels in the query region and that are the most discriminant. We call, for a given color c, the number of pixel L_{1c} and the discriminability level L_{2c} which is the percentage of pixel which are in the query region compared to rest of the pixels in the image of the same color. Each value of L_1 and L_2 are normalized as a percentage between 0 and 1, the score for each color is the sum of both term: $Score(c) = L_{1c} + L_{2c}$ (see fig. 3).

The discriminability level allows us to ignore colors that not only dominant in the query region but also in the rest of the image. For instance, black and white are most of the time present in the query but also in many contour and background regions which may create noise in the retrieval process.
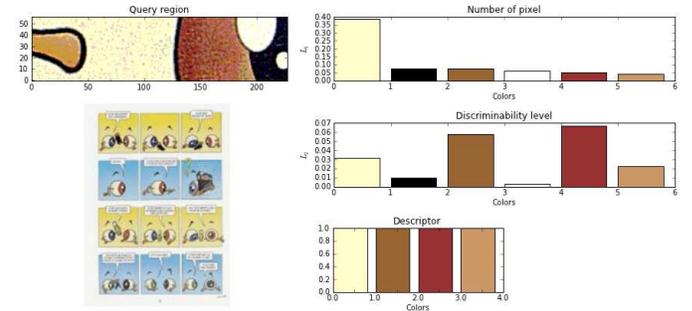

Fig. 3. The left part represents a query example after the color reduction process and its original image. Top-right part shows the color histogram of the 6 first colors with the highest number of pixel in the query as a percentage from the total number of pixel in the query (L_1). Middle-right figure represents the corresponding discriminability level for each color according the other pixel in the page. Bottom-right figure represents the corresponding query descriptor of size 4.

D. Object retrieval

Comics have a particular structure that allows different approaches for object retrieval. An interesting point is that similar objects often appear several times in each image (page) of the album.

First, for each page and color value (RGB) in the descriptor, we compute the corresponding color mask (see fig. 4). Second we find the object position in each pages by using a sliding window approach at different window size and for each mask. The window sizes are defined according to the user query size (maximum between the query width or height) which gives an information about the definition of the document $S = \{s_0, s_1, ..., s_n\}$ in order to retrieve the object at different scales.
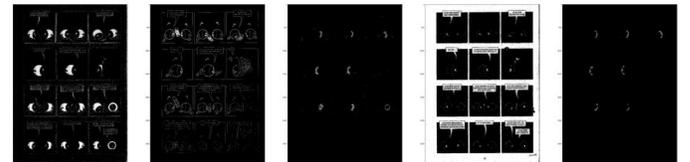

Fig. 4. Example of color mask correponding to the 5 first colors in the histogram of figure 3.

We define the detection confidence C according to the number of identical colors between the query descriptor D and each sliding window positions $W = \{w_0, w_1, ..., w_n\}$ (see eq. 1). This is equal to the cardinality of the intersection

between the two color set D and w_i. The confidence is maximal when all the colors in the query descriptor are contained by the window (high probability of the object to be at this location).

$$C(w_i) = \frac{|D|}{|D \cap w_i|} \quad (1)$$

where C is the confidence value for the region w_i, $|D|$ the cardinality of the color set D and $|D \cap w_i|$ the cardinality (number of element) of the intersection between D and w_i.

The confidence value is compared to a threshold value T to decide whether or not the detection is correct. By using a multi-scale approach, a same region may be detected at different scales. In order to keep only the smallest region that include a minimal amount of background information, we compute the percentage of pixel p which are of a color from the descriptor for each detection region. A post processing step removes multiple object detections by keeping only the regions (windows) that are not overlapped by other regions with a higher p value (best detected). See algorithm 1 and figure 5.

Algorithm 1 Object retrieval

//Detection
for each s in S **do**
    for each w in W **do**
        if $C(w) >= T$ **then**
            add w to the detected region list R
        end if
    end for
end for

//Filtering
for each r in R **do**
    if r does not overlap better detected region **then**
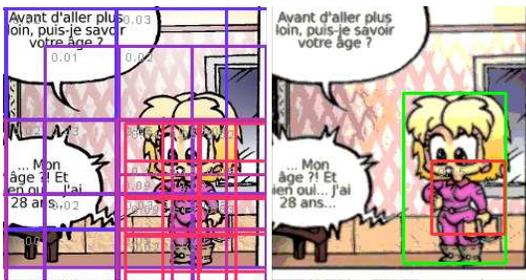 r is a candidate region
    end if
end for

Fig. 5. Multi-scale detection example. The left image shows the detections at different scales, the coloration depends on to the p value, blue to red for low to high. The right side image shows the remaining best detection after the filtering step.

III. EXPERIMENTS

In this section we evaluate the proposed method on different comics character from different albums. Examples of comics images we used are available in the eBDtheque dataset [12]. All the materials for reproducing and comparing the results presented in this paper are publicly available through http://ebdtheque.univ-lr.fr/references/.

A. Experimental setup

To perform the experiments, we asked the users to quickly highlight the object position as described in section II-A. The selection usually includes hair, head and top body colors because lower body parts are often hidden by the frame border or the posture. From the user pixel selection, we computed the bounding box to show that the system does not need a precise object color selection. We reduced the number of color to a standard web palette P_c of 256 colors (from the Python Image Library) because comics are colorful drawings similar to web pages and icons that contain a limited number of color. We fixed the descriptor size to $N = 6$ according to the evaluation figure 6. In the retrieval process, we set $T = N$ which means that the confidence of the candidate region equal 100% (all the query colors must be present in the candidate region to be considered as a correct detection). We used three squared window sizes according to the query height: half of the query height, query height and twice the query height.
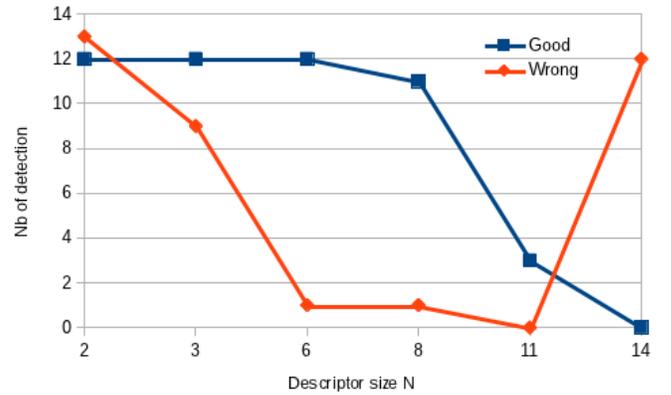

Fig. 6. Number of correct (blue square) and wrong (red diamond) detection for different descriptor sizes for 11 pages from a testing album. The highest difference is for $N = 6$.

B. Performance evaluation

The following results were obtained with the common evaluation measures of recall, precision at object level which means that we considered as correctly detected the objects that are overlapped by a detected region without considering the percentage of overlapping. Recall (R) is the number of correctly detected object divided by the number of objects to detect. Precision (P) is the number of correctly detected objects divided by the number of detected objects. We consider as candidate regions, regions that have $C = 100\%$ confidence (contains at least all the descriptor colors).

We evaluated our method on 34 comics pages from 10 different albums which represent 22 comic characters that appear 352 times in total. We obtained 90.3% recall and 46.7% precision in average. Results are detailed in table I where album 1 to 10 correspond to MIDAM GAMEOVER T05, TARZAN

COCHON, CYB COSMOZONE, CYB MAGICIENLOOSE, CYB MOUETTEMAN, LAMISSEB LESNOEILS1, LUBBIN LESBULLESDULABO, MIDAM KIDPADDLE7, PIKE BOYLOVEGIRLS T41 015 and TRONDHEIM LES TROIS CHEMINS 005 respectively from the eBDtheque dataset [12].

TABLE I.   CHARACTER DETECTION PERFORMANCE

Album	Nb pages	Nb character	Character ID	Recall	Precision
1	10	2	0	92.3%	16.8%
			1	91.8%	65.0%
2	1	1	2	75.0%	26.5%
3	5	4	3	100.0%	50.0%
			4	93.3%	60.9%
			5	100.0%	33.3%
			6	100.0%	100.0%
4	1	1	7	85.7%	85.7%
5	4	2	8	100.0%	75.0%
			9	50.0%	27.8%
6	5	2	10	71.8%	10.1%
			11	100.0%	29.9%
7	2	2	12	100.0%	13.8%
			13	100.0%	54.5%
			14	100.0%	55.6%
8	2	2	15	81.8%	75.0%
			16	100.0%	12.7%
9	2	1	17	83.3%	58.8%
10	2	4	18	81.3%	35.1%
			19	100.0%	48.4%
			20	92.9%	17.1%
			21	90.9%	24.4%
10	**34**	**22**	**22**	**90.3%**	**46.7%**

C. Detection results

We evaluated present interesting detection results based on a six colors user defined query descriptor (see section II). The processing time was about 6 seconds per image for a A4 300DPI image on a regular machine. Most of the time consumption is for the mask creation and it is proportional to the number of mask to create. Detection result examples are presented figure 7.

The detected region aims to localize the smallest regions containing all the colors of the query descriptor. This region is usually smaller than the ground truth region which is defined at character bounding box level. A post processing is needed to compute the character bounding box from the detected region by extending it to the color region boundaries.

Correct detections show the variety of comics character position and deformation that we are able to detect with the presented framework. There are few missing detection and over detection are essentially due to other comics character detection and image pre processing (color reduction).

IV. DISCUSSION

The proposed framework gives promising results for contemporary comics, we also tried on historical such as FRED PHILEMON12 and MCCALL ROBINHOOD from the eBDtheque dataset but the printing process generates a lot of noise (thickness of ink).A pre-processing smoothing [13] is required in this case. The color palette we used has 256 colors, we believe that we can improve the presented method by computing the palette from the user's query in order to ignore unwanted colors and speed up the process. The sliding windows approach does not allow to segment comics character at a pixel level, a distance measure between the color mask

Fig. 7.   Each line shows a character query region bounding box in white (left column), a correct detection (middle column) and a wrong detection (right column). Green rectangles represents the ground truth region and the red rectangle corresponds to the region detections.

regions could solve this. Also, once we retrieved all color similar objects, we could learn a shape model and try to find more objects occurrences based on shape information in a second stage, this will be a future work.

V. CONCLUSION

We have proposed and evaluated a new comics character detection framework based on a simple user interaction. From

one single object color example, we have shown that the presented framework is able to retrieve most of the character representation independently from posture and scale. The proposed approach relies on the discriminability of the query colors, further effort has to be made to retrieve similar comics character that have slightly different color due to shadow or scanning device.

ACKNOWLEDGMENT

REFERENCES

[1] W. Sun and K. Kise, "Similar partial copy detection of line drawings using a cascade classifier and feature matching." in *ICWF*, ser. Lecture Notes in Computer Science, H. Sako, K. Franke, and S. Saitoh, Eds., vol. 6540. Springer, 2010, pp. 126–137.

[2] W. Sun and K.Kise, "Detection of exact and similar partial copies for copyright protection of manga," *International Journal on Document Analysis and Recognition (IJDAR)*, pp. 1–19, 2013. [Online]. Available: http://dx.doi.org/10.1007/s10032-013-0199-y

[3] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[4] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *International Conference on Computer Vision & Pattern Recognition*, C. Schmid, S. Soatto, and C. Tomasi, Eds., vol. 2, INRIA Rhône-Alpes, ZIRST-655, av. de l'Europe, Montbonnot-38334, June 2005, pp. 886–893. [Online]. Available: http://lear.inrialpes.fr/pubs/2005/DT05

[5] F. S. Khan, M. A. Rao, J. van de Weijer, A. D. Bagdanov, M. Vanrell, and A. Lopez, "Color attributes for object detection," in *Twenty-Fifth IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2012)*, 2012. [Online]. Available: http://www.cat.uab.cat/Public/Publications/2012/SRV2012

[6] M. Wang, R. Hong, X.-T. Yuan, S. Yan, and T.-S. Chua, "Movie2comics: Towards a lively video content presentation." *IEEE Transactions on Multimedia*, vol. 14, no. 3-2, pp. 858–870, 2012. [Online]. Available: http://dblp.uni-trier.de/db/journals/tmm/tmm14.html#WangHYYC12

[7] H. N. Ho, C. Rigaud, J.-C. Burie, and J.-M. Ogier, "Redundant structure detection in attributed adjacency graphs for character detection in comics books," in *Proceedings of the 10th IAPR International Workshop on Graphics Recognition (GREC)*, Bethlehem, PA, USA, 2013.

[8] A.-P. TA, "Inexact graph matching techniques: Application to object detection and human action recognition," Ph.D. dissertation, INSA de Lyon, 20, rue Albert Einstein, 69621 Villeurbanne Cedex, France, 2010.

[9] W. Sun, K. Kise, J.-C. Burie, and J.-M. Ogier, "Specific comic character detection using local feature matching," in *Proceedings of International Conference on Document Analysis and Recognition (ICDAR 2013)*, Washington, USA, 2013.

[10] C. Rigaud, D. Karatzas, J. Van de Weijer, J.-C. Burie, and J.-M. Ogier, "Automatic text localisation in scanned comic books," in *Proceedings of the 8th International Conference on Computer Vision Theory and Applications (VISAPP)*. SCITEPRESS Digital Library, 2013.

[11] C. Rigaud, D. Karatzas, J. van de Weijer, J.-C. Burie, and J.-M. Ogier, "An active contour model for speech balloon detection in comics," in *Proceedings of International Conference on Document Analysis and Recognition (ICDAR 2013)*, Washington, USA, 2013. [Online]. Available: http://www.cat.uab.cat/Public/Publications/2013/RKV2013a

[12] C. Guérin, C. Rigaud, A. Mercier, and al., "ebdtheque: a representative database of comics," in *Proceedings of International Conference on Document Analysis and Recognition (ICDAR)*, Washington DC, 2013.

[13] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via l0 gradient minimization," *ACM Transactions on Graphics (SIGGRAPH Asia)*, 2011.