Dominant Color Palette Extraction in Resumes using the New Color Pixel Quantifier Algorithm

N. N. Perera
Faculty of Information Technology
University of Moratuwa
Moratuwa, Sri Lanka
nimnaperera98@gmail.com

S. P. Warusawithana
Faculty of Information Technology
University of Moratuwa
Moratuwa, Sri Lanka
supunawa@gmail.com

R. L. Weerasinghe
Faculty of Information Technology
University of Moratuwa
Moratuwa, Sri Lanka
reshakalakshan@gmail.com

T. M. Hindakaraldeniya
Faculty of Information Technology
University of Moratuwa
Moratuwa, Sri Lanka
tharushi.18@itfac.mrt.ac.lk

G. U. Ganegoda
Faculty of Information Technology
University of Moratuwa
Moratuwa, Sri Lanka
upekshag@uom.lk

***Abstract*—In the realm of resume analysis and enhancement, the extraction of dominant color palettes plays a pivotal role in assessing the visual impact of resumes. Existing methods designed for images with extensive color ranges have proven to be suboptimal when applied to the distinct context of resumes, which inherently possess a limited color palette. This paper introduces a novel approach that addresses this challenge effectively and efficiently. By minimizing the time required for palette extraction without compromising accuracy, the proposed method offers a practical solution for resume feedback systems. It is important to clarify that this research neither rejects nor supports existing methods; instead, it presents an alternative, tailor-made solution for resume analysis. In summary, this paper sets a promising precedent for more streamlined and functional dominant color palette extraction methods in the context of resumes, promising advancements in resume analysis and improvement.**

Keywords— color palette extraction, resume design analysis, color pixel quantifier

I. Introduction

A resume is a type of electronic document that is used as the key document of an individual during his/her time of pursuing new career opportunities. A typical resume would contain the following details but are not limited to education, technical skills, non-technical skills, and the working history of that particular individual [1], [2]. Even though textual content about the individual is the core of the resume [3] when it comes to getting the first impression of the recruiters, the design of the resume plays a vital role. Hence, when an individual is building their resume, they should focus not only on the content of the resume but also on the design such that it looks impressive and attractive to the recruiter while being professional [4].

When considering the design of a resume there are a few factors to be considered, such as originality, easily understandable format and layout, and last but not least color combination usage [1], [5]–[7]. Having a good color combination usage in the resume would lead to a good first impression on the resume which leaves positive feedback about the person. But to analyze the color combinations used, first, it needs to extract the dominant color palette used in the resume.

But when considering the currently implemented methods in [8]–[12] for this purpose, they are not perfectly aligned and do not give the results as expected. Currently implemented methods are designed for extracting dominant color palettes in images, which use thousands to millions of colors in a single frame, by using unsupervised clustering techniques such as KMeans or dynamic clustering algorithms such as the MeanShift algorithm [13]. But in the context of a resume, it is not the case. Resumes only use a limited number of colors in the document. Even though, when considering a higher number of color values in an image, clustering is considered the best method to extract the dominant color palette, in the context of a resume, it is not considered as much as efficient, since it takes a considerable amount of time to extract the color palette of the resume, despite having a limited number of colors.

Therefore, bridging this gap, in this paper, a new algorithm has been proposed for extracting the dominant color palettes in resumes, by using a color pixel quantifying mechanism. It will be shown that the newly proposed algorithm reduces the computational cost which ultimately leads to a lesser amount of time for the execution. Further, the paper will benchmark the currently implemented methods against the newly proposed method, which signifies the efficiency and effectiveness of accomplishing the given task.

By proposing this tailored approach that optimizes efficiency without sacrificing precision, this research presents a practical solution for extracting color palettes in documents with limited color ranges. The findings not only offer immediate applicability in resume feedback systems but also open avenues for rethinking color extraction in various document types. The nuanced discussion on the limitations of existing methods and the proposed solution encourages a paradigm shift in the way dominant color palettes are approached, marking a substantial step forward in document analysis and design evaluation.

The rest of the paper is organized as follows. In Section 2 the method of the newly proposed algorithm will be discussed. Section 3 will depict the evaluation of the proposed method including the comparison between current methods and the newly proposed algorithm, emphasizing its efficiency and effectiveness, following the discussion and conclusion of the paper in Sections 4 and 5 respectively.

II. Methodology

This research introduces a groundbreaking algorithm for extracting dominant color palettes in documents, focusing on the unique context of resumes. Conventional methods,

979-8-3503-5950-3/23/$31.00 ©2023 IEEE

designed for images with extensive color ranges, prove inefficient for limited-color documents like resumes. The proposed algorithm employs a specialized color pixel quantifying mechanism, addressing the limitations of existing methods. This innovative approach significantly enhances efficiency in extracting dominant color palettes, setting a new standard for document analysis methodologies.

When contemplating the extraction of the dominant color palette in the proposed method, there are three main steps, which can be identified as the preprocessing step, color occurrence quantization, and similarity measurement between quantized colors. In this section, how the above-mentioned steps have been carried out, will be explained.

A. Preprocessing

For the preprocessing step of the algorithm, first, it is needed to convert the resume from the Portable Document Format (PDF) to an image format for the image manipulation tasks. After the conversion, to use only lower computational power for further processing, the image of the resume was resized to a smaller size while keeping the color features of the image.

B. Color Occurrence Quantization

In order to extract the dominant color palette there are two main things to be done.

1. Differentiate dominant colors from non-dominant colors.
2. Selecting the best representation for the palette from out of all dominant colors.

In the color quantization stage, the first task of differentiating dominant colors from non-dominant colors is achieved.

For that purpose, in the previously explored studies, the researchers have used color clustering using the K-Means algorithm [8]. According to later studies [9], [10], these color clustering approaches have been improved using dynamic clustering techniques. However, the main drawback of color clustering is that at a higher number of clusters, it takes too much time for clustering. Further random initialization of cluster center points can make it worse.

But in this newly proposed approach for color occurrence quantization, a very simple method of quantization has been employed, which is traversing through each pixel. Even though the time taken to traverse may be changed due to the size of the image, the resizing of the image preview of the resume, which has been done in the preprocessing step, ensured that it does not take much more time for the traverse. While traversing through the resized image preview of the resume, a map of color and its number of occurrences has been kept in the memory for further processing.

Despite requiring a greater number of colors for the palette, it achieves this with consistent time efficiency and reduced variance. This advantage stems from the fact that the mentioned color map comprehensively encompasses all instances of unique color values within the memory. In that way, the new approach effectively tackles the previously mentioned challenge.

C. Similarity measurement

When comes to the similarity measurement stage, it is very crucial because otherwise, selecting dominant colors with a higher similarity can leave the best representational colors at rest. Therefore, to select the best color palette, it is needed to use a distance measurement formula, which can be adopted in measuring color distance in RGB space.

In this proposed approach, for the dissimilarity measurement, Euclidian distance has been used. The Euclidean distance is a fundamental metric in spatial analysis and serves to compute the straight-line distance between two points in a multi-dimensional space. In the given context, each color can be represented as a point in a three-dimensional RGB (Red, Green, Blue) space.

Given a color C and color D from the document, both defined in RGB, the Euclidean distance (ED) between them is mathematically expressed as in (1):

$$ED(C,D) = \sqrt{(R_C - R_D)^2 + (G_C - G_D)^2 + (B_C - B_D)^2} \quad (1)$$

Where R_C, G_C, B_C denote the red, green, and blue components of color C, respectively, and R_D, G_D, B_D represent the corresponding components of color D. A smaller Euclidean distance implies a higher similarity between the colors, so those are discarded from taking as a distinct color for the palette. Further by running the similarity measure from most occurred to least occurred color pixel values, the most representative color palette has been selected as the final output.

III. Evaluation

The main purpose of proposing this new method was to get a more accurate and reliable dominant color palette output for resumes while showing good performance in the proposed method as well. By considering these two main factors, the evaluation of this proposed method has been conducted in two main methods of evaluation.

1. Objective evaluation.
2. Subjective evaluation.

The goal of conducting an evaluation in the hybrid method is to evaluate the proposed method(s) more reliably while getting statistical-based and behavioral-based evaluations for the final assessment [14].

Under the objective evaluation, it is supposed to get a statistical value for the evaluation of the proposed method. When considering the proposed method, since it was supposed to perform efficiently for the given task, the time taken to extract the dominant color palette from resumes has been used as an objective evaluation. For the objective evaluation, ten unseen resumes have been selected from the resume sample, and those ten resumes have been used for the dominant color palette extraction task. Using the newly proposed method, and the K-means color clustering extraction, dominant color palettes were extracted while recording the time taken to extract the dominant color palette. While doing this experiment, the extraction of the color palette was conducted by extracting two, three, four, and five colors for the palette, respectively. After the experiment, the time taken to extract the number of colors for the palette has been averaged for ten resumes. From that, the average time taken to extract the designated number of colors for the dominant color palette has been used as the final statistic.

The graph in Fig. 1 depicts the average time taken in seconds to extract two, three, four, and five colors for the dominant color palette using K-Means color clustering extraction and the newly proposed color pixel quantifier algorithm. The blue line depicts the average time taken by K-Means color clustering extraction and the orange line depicts the average time taken by the newly proposed color pixel quantifier algorithm.

As the subjective evaluation, twenty-one human subjects were utilized for rating the accuracy of the output which has been generated by the newly proposed method. This has been conducted through a survey using the same ten resumes used in the objective evaluation. The survey's design was meticulously crafted to elicit comprehensive feedback on various aspects of the evaluated resumes. A diverse pool of participants was selected through a purposive sampling method, ensuring representation across demographics and professional backgrounds. The selection process prioritized individuals with experience in recruitment, human subjects experienced in the color theory such as artists, and neutral human subjects, enhancing the relevance of the feedback to industry standards and also the aesthetic value. Rigorous standards were employed to determine rating accuracy, incorporating predefined criteria based on established resume evaluation metrics. Participants were provided with clear guidelines and calibrated rating scales, minimizing subjectivity and enhancing the consistency of responses.

Following the above standards, in the first half of the survey, subjects were tested for color blindness. The color blindness for eight major colors has been assessed during this section. This step holds paramount importance due to the inherently color-sensitive nature of the subsequent assessments. Ensuring that participants possess accurate color perception is instrumental in preventing skewed or invalid results. By systematically addressing and confirming participants' color visual acuity, this approach fortified the reliability and validity of the ensuing subjective evaluations, reinforcing the integrity of the study's findings. In the second half of the survey, subjects were requested to rate the ten resumes one by one for how much the dominant color palette extracted by the proposed method matches the resume's dominant color palette, according to their preference and eyesight. Fig. 2 to Fig. 5 depicts four results out of eight, which were gathered regarding the color blindness of the subjects. Fig. 6 to Fig. 8 depicts three results of the ratings out of ten resumes that have been rated by the human subjects.

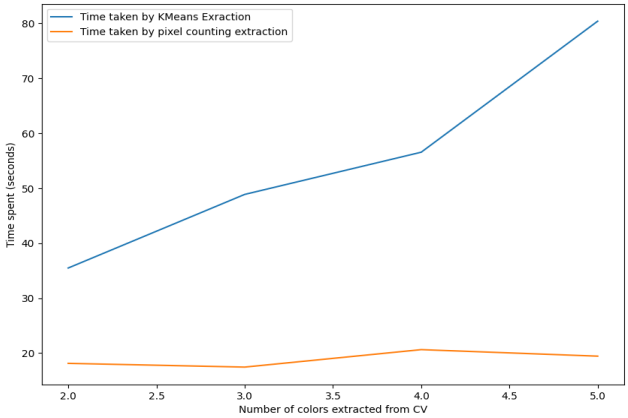

Fig. 1. The average time taken in seconds to extract two, three, four and five colors for the dominant color palette using K-Means color clustering extraction and the newly proposed color pixel quantifier algorithm.

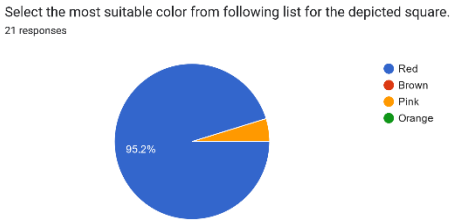

Fig. 2. The result of the red color blindness of the subjects

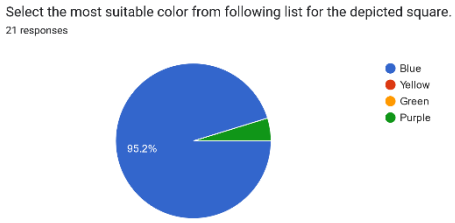

Fig. 3. The result of the blue color blindness of the subjects.

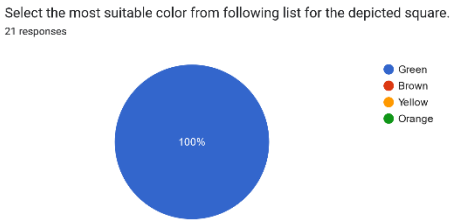

Fig. 4. The result of the green color blindness of the subjects

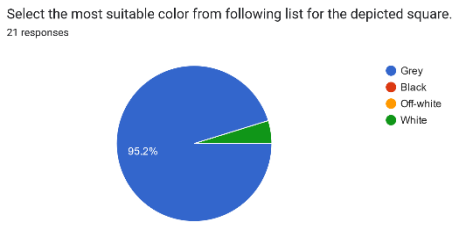

Fig. 5. The result of the grey color blindness of the subjects.

The summary of the ratings which have been given by the twenty-one human subjects according to their perspective is shown in Table 1. In the table, rating one indicates that the extracted dominant color palette poorly matches with original color palette of the resume, and rating 5 indicates that the extracted dominant color palette perfectly matches with original color palette of the resume.

IV. Discussion

In this section, we delve into the ramifications of our findings, considering their alignment with existing literature and the potential for future research directions. We also explore the limitations encountered during our study and their potential impact on the validity of our results. Through this comprehensive analysis, we aim to distill the essence of our research, and the broader context of our work.

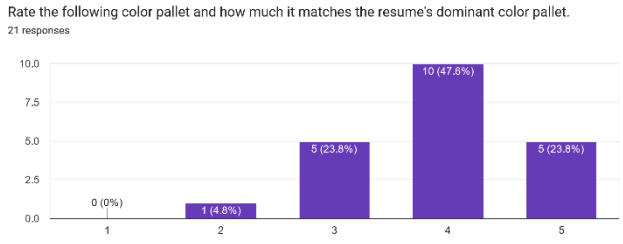

Fig. 6. The results of a resume that has been rated by human subjects (i).

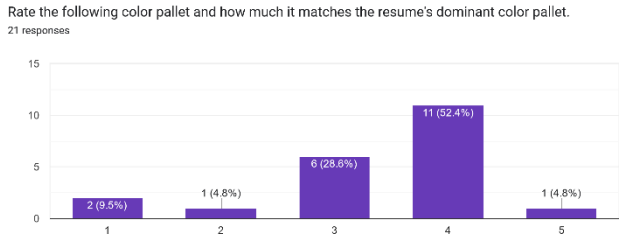

Fig. 7. The results of a resume that has been rated by human subjects (ii).

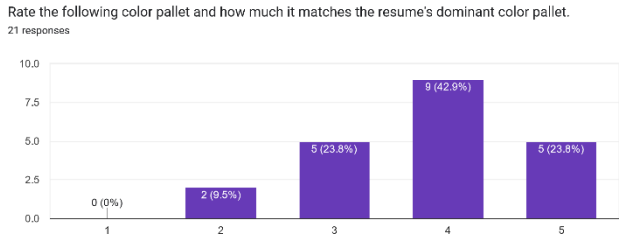

Fig. 8. The results of a resume that has been rated by human subjects (iii).

When considering the above-mentioned objective evaluation, it is notable in Fig. 1 that the newly proposed method is very efficient when compared to the existing K-means color clustering approach. In the color clustering approach, when the number of extracted color counts for the palette increases, it will increase the average time taken for the extraction exponentially. But unlike that, the newly proposed method spends an equal amount of time for the extraction task, signifying its linear time distribution.

As shown in Fig. 2 to Fig.5 it is notable that out of all twenty-one human subjects which were utilized in the subjective evaluation, one person has been shown color blindness for blue and red colors. Hence the responses of that individual have been excluded to get the final exact evaluation result set. After adjusting the results of the ratings, using the weighted average, a set of single statistics for individual resumes and rating categories have been derived to represent the effectiveness of the proposed method. When deriving the statistic, the number of subjects who rated the same rating was used as the weightage for each rating value. Table 2 depicts the adjusted result set and the derived statistics for each rating.

As shown in the derived statistics, eight out of ten resumes were able to gain an equal or more than average rating of 3.5, which is equivalent to more than 70% match of the dominant color palette extracted by the proposed method and original palette of the resume. When taking the perspective of the human subjects regarding the extracted dominant color palettes, out of twenty subjects 82% considered the extracted dominant color palettes worth a rating which is equal to or greater than three. Out of those 82%, 72% of subjects have given a rating that is equal to or greater than 4. These two types of evaluation are enough to prove that the newly proposed method shows considerable effectiveness and efficiency for the designated purpose.

As limitations and further work of the proposed method, the main limitation which has been identified is the proposed method only implemented for the resumes which are in Portable Document Format (PDF). When considering the real-world scenario there can be resumes prepared in document formats as well. Hence, it would be very accommodating if the proposed method could be modified to assist those document types as well. Further, for the similarity measurement in the proposed method, since it uses a predetermined value for identifying the margin, it can be modified to use a statistical approach to determine the margin value.

V. Conclusion

In conclusion, this paper addressed a significant challenge in the field of dominant color palette extraction for resumes.

TABLE I. The Summary of the Ratings which has been Given by the Twenty-one Human Subjects

Resume	Number of people who given the rating				
	1	***2***	***3***	***4***	***5***
Resume 01	0	3	4	11	3
Resume 02	0	5	2	11	3
Resume 03	9	7	4	1	0
Resume 04	0	3	2	11	5
Resume 05	0	1	7	11	2
Resume 06	0	1	5	10	5
Resume 07	2	1	6	11	1
Resume 08	0	2	5	9	5
Resume 09	1	1	9	6	4
Resume 10	1	1	5	8	6

TABLE II. The Adjusted Result Set and the Derived Stistics for Each Rating

Resume	Number of people who given the rating					Average rating
	1	***2***	***3***	***4***	***5***	
Resume 01	0	2	4	11	3	**3.75**
Resume 02	0	5	2	10	3	**3.55**
Resume 03	9	6	4	1	0	**1.85**
Resume 04	0	3	2	10	5	**3.85**
Resume 05	0	1	6	11	2	**3.70**
Resume 06	0	1	5	9	5	**3.90**
Resume 07	2	1	5	11	1	**3.40**
Resume 08	0	2	4	9	5	**3.85**
Resume 09	1	1	8	6	4	**3.55**
Resume 10	1	1	5	8	5	**3.75**
	1.3	**2.3**	**4.5**	**8.6**	**3.3**	

The currently implemented methods, despite their proficiency in handling images with an extensive color range, prove inefficient when applied to the context of resumes, which inherently employ a limited color palette. To bridge this gap, we introduced a novel and more efficient approach. Our proposed method not only alleviates the time-consuming nature of existing techniques but also exhibits superior effectiveness when dealing with the distinctive color requirements of resumes.

It is important to note that our research neither supports nor rejects existing methods; rather, it puts forth an alternative solution tailored specifically for resumes. Our innovation stems from the growing need for extracting dominant color palettes in resumes, a crucial step in analyzing color combinations for enhancing resume feedback systems.

Nevertheless, it is imperative to acknowledge the limitations of our approach, as discussed in the preceding sections. These limitations should serve as valuable pointers for future research in this domain. In summary, this work paves the way for more efficient and practical methods of dominant color palette extraction in resumes, offering substantial benefits to the field of resume analysis and enhancement.

ACKNOWLEDGMENT

This research study is supported by the Faculty of Information Technology, University of Moratuwa, Sri Lanka under the supervision of the Department of Information Technology.

REFERENCES

[1] R. M. Schramm and R. Neil Dortch, “An analysis of effective resume content, format, and appearance based on college recruiter perceptions,” *Bull. Assoc. Bus. Commun.*, vol. 54, no. 3, pp. 18–23, 1991.

[2] “7 Simple But Effective Ways to Make Your CV Stand Out | Top Universities.” https://www.topuniversities.com/blog/7-simple-effective-ways-make-your-cv-stand-out (accessed Jul. 22, 2022).

[3] N. H. Anderson and A. A. Barrios, “Primacy effects in personality impression formation.,” *J. Abnorm. Soc. Psychol.*, vol. 63, no. 2, p. 346, 1961.

[4] E. Furtmueller, C. Wilderom, and R. M. Mueller, “Online resumes: Optimizing design to service recruiters,” 2010.

[5] J. K. Arnulf, L. Tegner, and Ø. Larssen, “Impression making by résumé layout: Its impact on the probability of being shortlisted,” *Eur. J. Work Organ. Psychol.*, vol. 19, no. 2, pp. 221–230, 2010.

[6] A. M. Mithun, Z. A. Bakar, and W. M. S. Yafooz, “The impact of web contents color contrast on human psychology in the lens of HCI,” *Int. J. Inf. Technol. Comput. Sci.*, vol. 11, no. 10, pp. 27–33, 2019.

[7] R. H. Hall and P. Hanna, “The impact of web page text-background colour combinations on readability, retention, aesthetics and behavioural intention,” *Behav. \& Inf. Technol.*, vol. 23, no. 3, pp. 183–195, 2004.

[8] I. Pavan Kumar, V. P. Hara Gopal, S. Ramasubbareddy, S. Nalluri, and K. Govinda, “Dominant color palette extraction by K-means clustering algorithm and reconstruction of image,” in *data engineering and communication technology*, Springer, 2020, pp. 921–929.

[9] Y. Chang and N. Mukai, “Color Feature Based Dominant Color Extraction,” *IEEE Access*, vol. 10, pp. 93055–93061, 2022.

[10] N.-C. Yang, W.-H. Chang, C.-M. Kuo, and T.-H. Li, “A fast MPEG-7 dominant color extraction with new similarity measure for image retrieval,” *J. Vis. Commun. Image Represent.*, vol. 19, no. 2, pp. 92–105, 2008.

[11] S. Kiranyaz, S. Uhlmann, T. Ince, and M. Gabbouj, “Perceptual dominant color extraction by multidimensional Particle Swarm Optimization,” *EURASIP J. Adv. Signal Process.*, vol. 2009, pp. 1–13, 2010.

[12] H. Shao, Y. Wu, W. Cui, and J. Zhang, “Image retrieval based on MPEG-7 dominant color descriptor,” in *2008 The 9th international conference for young computer scientists*, 2008, pp. 753–757.

[13] M. A. Carreira-Perpinán, “A review of mean-shift algorithms for clustering,” *arXiv Prepr. arXiv1503.00687*, 2015.

[14] C. Delgorge-Rosenberger and C. Rosenberger, “Evaluation of medical image compression,” *Handb. Res. Adv. Tech. Diagnostic Imaging Biomed. Appl.*, pp. 447–459, 2009, doi: 10.4018/978-1-60566-314-2.CH028.