Classification applications in Computer Vision with human faces

Sebastian Velasquez
s1700260

The face is perhaps the part of the human body that carries the majority of visual information about a person. Several systems have exploited the data from faces to solve a variety of computer vision tasks. In this review, I present and compare several types of information from faces in 2D images and how several applications use them to accomplish several tasks. Later, I provide an analysis of three systems that leverage the properties of those types of information to address different problems, and how they could be improved or used for other purposes. Finally, I provide conclusions and further work.

1 Introduction

The range of applications in computer vision that use human faces varies from entertainment to surveillance including health, sports, fashion and more. A significant number of these systems belong to the field of classification being the most relevant tasks detection, recognition, and verification. Depending on the domain, some approaches address a given task as simple as using a single feature of a subject whereas others use more complex analysis that includes correlation between different sets of attributes.

The variability in the amount of data that every alternative use supposes that the human face is a source of vast and flexible information. Depending on the features to analyze and the

way the systems analyze them, the results can be dramatically different. For this reason, it is essential to have a good understanding of the domain of the application to get proper results. For instance, in a system that is meant to determine the mood of a person, it makes more sense to use the facial expressions rather than the hair or eye color.

For the common eye, it is easy to identify a set of features from almost any object (faces included). However, in computer vision, this is a non-trivial task. Several general purpose feature detectors from images have been developed e.g. (*1*) and (*2*), as well as domain specific ones: (*3*) and (*4*). No matter their domain, the ultimate goal of these approaches is to provide information in a meaningful way for a system. This way, an application can perform a given task with good results. For classification, some alternatives can extract features and categorize faces at once (*5*).

The information extracted from faces can be used to feed a variety of systems for different purposes. Many of the alternatives use machine learning methods, being neural networks a popular choice for tasks like face detection (*6*), face recognition (*7*) and face verification (*8*); all of them intrinsically classification tasks.

2   Face information

Even though several approaches address the problem of feature detection and they have proved to give good results, it is essential to understand the types of information that we can get from faces. This way it can be possible to get more insights about how some systems leverage the data from a human face and how they can be improved or modified to perform other tasks. Here, I will focus on the geometry and color information provided by faces.

2.1 Geometry

Geometry-wise, there are three critical characteristics of the human face: structure, vertical symmetry, and components variability. In regards to the structure, the face has a well-defined composition which means that the positions of its different components have small variance among various subjects, e.g., the mouth is close to the bottom, the eyes are close to the top, etcetera. Some applications take advantage of this property for synthesizing data to train neural networks (*9*), copying a makeup style from one face to another (*10*), or merely replacing faces in photographies for entertainment purposes (*11*). Moreover, the well-defined structure of a face along with its enormous differences to other elements make possible systems for face detection (*12*).

The vertical symmetry of the human face is another salient geometric characteristic that several applications exploit. In the task of face recognition, this property has been used to reduce the effects of illumination in face recognition systems (*13*). The main drawback of this property is its sensitivity to pose variations and facial expressions, an issue that dramatically reduces the performance in face recognition approaches. Several alternatives have tried to minimize these effects by using neural networks (*14*) and probabilistic models (*15*).

Finally, the variability of the face components from subject to subject (shape, relative position, relative size, etcetera) is high enough that it is possible to have approaches for face recognition (*16*), a task that leads to systems for authentication or re-identification. Also, the components of a face in a single subject are not static (their shapes can vary) which is evident in facial expressions. While this can be a problem for specific tasks, some applications leverage this property to recognize moods or emotions (*17*). Some systems use this characteristic to facilitate human-computer interaction (*18*).

Other than facial expressions, subjects can modify the geometry of their faces by using makeup, wearing accessories like glasses and earrings, or even by changing their hairstyle.

This situation evidently increases the variability of a single face, which supposes a challenge for systems that rely on the geometric information of faces. Some alternatives have addressed this problem by performing the correlation between the hard face and the modified one (*19*). Other approaches have focused their efforts in the detection of such accessories (*20*) and (*21*).

2.2 Color

In computer vision, one of the most salient characteristics of color is the ability to represent it in a multitude of formats. Thanks to this property, it is possible to convert the color information in a picture from the default color space (usually represented with the red, green and blue components) to a more suitable format depending on the application domain. This attribute is leveraged by several systems for different purposes including explicit image detection (*22*) and face detection (*23*). Other applications that take advantage of color information include face tracking (*24*) and gaze tracking (*25*).

Several types of research have widely studied the human skin color in digital images over the past years. One of the significant advances made in previous investigations in this area is the determination of the proper color spaces to study the human skin. Even though an appropriate analysis of almost any color format can lead to good results in applications to detect skin regions (*26*), color spaces that separate the chroma and the luminance components (YCbCr) have demonstrated to provide the best results. The main reason behind these outcomes is that the luminance element (Y attribute) does not give much information about the skin and many systems discard it as stated in (*27*).

Even though some skin detection applications dismiss the luminance information; other alternatives need this information to discriminate faces from other human body parts because it provides many clues about the human face. However, the main drawback in 2D images is the sensitivity of color to illumination changes which is a major issue that reduces the performance

of several systems. The main reason for such operation degradation is the use of images taken under well-controlled illumination conditions, something that is not always possible in the real world. Some approaches try to solve this problem by using Gaussian filters and performing Principal Component Analysis (PCA) (*28*) or increasing the contrast of images and filtering the best features from a face (*29*).

The skin is not the only source of color, other face components like the eyes and hair provide valuable information. The analysis of the data from those elements has proved to be useful in applications like gaze tracking (*30*) and person recognition (*31*). One significant disadvantage related to eyes in color images is the red-eye effect caused mainly by poor light conditions and the interaction of flashlight from cameras with the choroid. This fact has been well studied, and some alternatives have been proposed to correct it (*32*).

Similar to geometric features, face color is prone to changes too. Here the differences can also depend on the behavior of the subjects, e.g., the use of makeup. However, most of the color variability in an image comes from the configuration of the device to capture the scene (exposure time, shutter speed, etc.), and the environment setup (lighting, shadows, etc.). Nonetheless, skin color variability not only brings problems, but it can also be leveraged to make health-related analysis (*33*).

3 Classification applications with faces

There is a wide range of applications that lay their foundations in the classification task using images with faces. Based on the geometry and color information mentioned earlier, I will describe and analyze three systems that use that kind of data. The first method uses geometry information represented as a histogram of gradients (HOG) for face recognition (*34*). The second system detect faces based on skin region detection using color transformations (*35*). Finally, the third approach uses both geometry and color information to determine the presence

of makeup in faces (*36*).

3.1   Face recognition using HOG

As stated by (*34*) HOG is a technique founded on the number of edge orientations occurrences within the local vicinity in an image. In their work, the authors propose a method based on the previous project by (*37*). The main differences with the previous work are the approaches to calculate HOG, a combination of classifiers and the databases used to ran the experiments. The different methods to perform HOG are aimed to reduce the possible dependency on the facial landmarks localization of previous results. A facial landmark is a salient feature of a face, in other words, a point of interest.

There are two proposed ways to perform HOG in a face image. The first one lies in the location of the facial landmarks as well. However, the critical point here is redundancy. Once the facial markers are localized, the system performs HOG over each one of them at different patch sizes. That means that the bigger the patch size, the higher the redundancy of the retrieved information by HOG as seen in Figure 2 in (*34*). In the second method, an image face merely is divided into non overlapped patches, then HOG is calculated in each section.

As the authors mention, the redundancy of the first method may cause over-fitting if the classifier does not have a mechanism for feature selection. On the other hand, the second alternative could prevent that situation due to its non-overlapping nature. However, some amount of redundancy may be necessary to avoid the other undesired effect in machine learning models: under-fitting. In fact, some HOG implementations can be configured to create a certain amount of redundancy by adding normalization over overlapped blocks of cells.

Another option that could help to increase the robustness of the feature descriptor is the combination of several types of features. Even though the authors experimented combinations of HOG with PCA and Latent Dirichlet Allocation (LDA), other methods like Local Binary

Patterns may give good results. In fact, an alternative that combines HOG and LBP has proved to perform well in adverse situations like occlusion as seen in (*38*).

The authors propose a method to make the recognition of a face based on the idea of weak classifiers, something that has proved to be right in several scenarios including the very same face recognition task (*39*). Another machine learning method that can perform well in this kind of problems is neural networks as demonstrated in (*40*). However, the success of a neural network is highly tied to a significant amount of data for the training process.

Finally, the datasets used to perform the experiments presented considerable differences from one to another. As seen in Figure 4 of their work, the variance of the facial landmarks is entirely different from one database to the other. This situation is not a minor issue since the performance of the method varies according to the dataset used. There is no much that can be done here since each dataset has their characteristics, some of them were even built for a specific purpose. However, the naive alternative is to combine all the databases to get a more robust classifier.

Even though the proposed method does not irrupt in the computer vision field, it was able to improve the results of its ancestor. The main point in this solution is the way to extract face features; it uses the same mechanism of the previous approach, but its execution is different. As seen in many scenarios, a proper definition of attributes can lead to expected results and verify the validity of a hypothesis, something that this work accomplished. Further work could be aimed to test the performance of the feature extraction methodology in other tasks like face detection.

3.2   Face detection in color images

The approach for face detection proposed by (*35*) is highly focused in the analysis of color in images. The solution is motivated by the lack of previous color-based alternatives that are

robust enough to detect skin colors in different light conditions and complex background. The proposed method relies on two main components. The first one is aimed to detect face candidates based on some color transformations performed in an image. The second element focuses on the validation of the face candidates by detecting facial features.

The authors are aware of the challenges related to the definition of skin tones in color images due to the lighting conditions, and the previous work in the field. Based on this knowledge, the solution starts by compensating the lighting condition by using a white reference area, which allows reducing the number of false positives. As demonstrated in other related works, YCbCr is the preferred color space when it comes to human skin detection. However, unlike other solutions where the skin pixels were detected by merely checking their values against a predefined range, the algorithm performs a non-linear transformation that helps to reduce the false positives as well.

Once the face candidate region is detected, it is analyzed to find facial features. This analysis relies on the distinctive chroma and luminance components of eyes and mouth. In summary, the solution creates an intensity map that highlights the locations of the eyes and mouth. Such intensity map is created through a sequence of a nonlinear transformation of the image components in the YCbCr color space. Once the three regions are detected, all the triangle zones are computed and filtered under a threshold to identify a face finally.

As described, this solution makes an intense analysis of the color in an image. However, due to the lack of semantic analysis, some anthropomorphic areas can be detected as human faces, e.g., toys. Further work could consider an analysis of the context in the image to reduce the false positives and improve results. Finally, since the approach analyzes the luminance in an image, it could be used to detect fake or mounting photos.

3.3 Automatic Facial Makeup Detection with Application in Face Recognition

Face recognition is a well-studied task; many geometry-based approaches have been proved to provide excellent results under certain conditions. However, in the real world, many variables can reduce the performance of such solutions. Makeup is one of those variables that challenges face recognition. The work proposed by (*36*) is founded over the lack of previous solutions that address the problem of makeup and given that wearing makeup is a common habit by a high percentage of the population.

In their work, they propose a method that combines a set of color, geometric and texture features aimed to detect the presence of makeup in a face. Moreover, the solution is extended to be used in face recognition. As stated in their work, wearing makeup changes the perceptual appearance of people by highlighting characteristics and reducing flaws. Several types of research have proved that this situation causes avast impact in the performance of some recognition systems. For this reason, it is necessary to redefine the set of features used to feed those solutions.

The main contribution of this work is the redefinition of the facial features to consider the changes in shape, color, and texture due to wearing makeup. The primary objective is to provide a set of descriptors that precisely determine the presence or absence of makeup, so a face recognition system can be appropriately tuned to maintain its performance under such conditions.

Unlike other works that make use of the YCbCr color space when it comes to human skin, here they use the HSV color space (Hue, Saturation, Value) that highlights makeup regions in the saturation channel. The solution proposes a definition of color descriptors by combining the local and global features. This approach is a good idea since it adds some redundancy but keeps fidelity over all images. Shape and texture descriptors are simple, but they still contribute to improving the robustness of the feature vector.

Their experiments proved to give good results; however since the work was highly focused on mid-age female image samples, further actions could create a more general solution including other segments of the population. Moreover, since the approach used almost perfectly front face images, additional work can also include samples with different degrees of pose variation.

4   Conclusions and further work

As seen, the human face is a source of vast information. The different types of data can be leveraged to address a multitude of computer vision tasks. Despite the fact that geometric and color information have their strengths and drawbacks, rather than being mutually exclusive alternatives, they can be combined to compensate their weaknesses and maximize the performance of an application.

One of the key aspects of systems that rely on the use of facial information is the way they represent a face. Depending on the domain of the application some features are more useful than others. But some solutions go beyond that; it is not just about the kind of information used to address a problem, the methods used to process and present it are very relevant to the success of an approach.

Also, it is evident that a given task can be addressed in different ways. Nonetheless, another essential component of every approach is its foundation in previous work, this way new solutions can go beyond previous proposals. In the case of applications with human faces, there is a lot of work done to analyze their information and make use of them. However, the development does not seem to stop, and every time new alternatives are proposed. Further reviews could be aimed to investigate what kind of information from human faces could be retrieved from sources other than 2D images, including Laser Imaging Detection and Ranging (LIDAR) and medical imagery, and the applications that can be created.

References and Notes

1. C. Harris, M. Stephens, *Alvey vision conference* (Manchester, UK, 1988), vol. 15, pp. 10–5244.

2. D. G. Lowe, *Computer vision, 1999. The proceedings of the seventh IEEE international conference on* (Ieee, 1999), vol. 2, pp. 1150–1157.

3. G. Chow, X. Li, *Pattern Recognition* **26**, 1739 (1993).

4. A. Colmenarez, B. Frey, T. S. Huang, *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on* (IEEE, 1999), vol. 1, pp. 657–661.

5. M. K. Fleming, G. W. Cottrell, *Neural Networks, 1990., 1990 IJCNN International Joint Conference on* (IEEE, 1990), pp. 65–70.

6. H. A. Rowley, S. Baluja, T. Kanade, *IEEE Transactions on pattern analysis and machine intelligence* **20**, 23 (1998).

7. S. Lawrence, C. L. Giles, A. C. Tsoi, A. D. Back, *IEEE transactions on neural networks* **8**, 98 (1997).

8. Y. Sun, X. Wang, X. Tang, *Proceedings of the IEEE International Conference on Computer Vision* (2013), pp. 1489–1496.

9. G. Hu, X. Peng, Y. Yang, T. Hospedales, J. Verbeek, *arXiv preprint arXiv:1603.06470* (2016).

10. D. Guo, T. Sim, *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on* (IEEE, 2009), pp. 73–79.

11. D. Bitouk, N. Kumar, S. Dhillon, P. Belhumeur, S. K. Nayar, *ACM Transactions on Graphics (TOG)* **27**, 39 (2008).

12. Q. Zhu, M.-C. Yeh, K.-T. Cheng, S. Avidan, *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on* (IEEE, 2006), vol. 2, pp. 1491–1498.

13. Y.-J. Song, Y.-G. Kim, U.-D. Chang, H. B. Kwon, *Pattern Recognition* **39**, 1542 (2006).

14. F. J. Huang, Z. Zhou, H.-J. Zhang, T. Chen, *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on* (IEEE, 2000), pp. 245–250.

15. T. Kanade, A. Yamada, *Computational Intelligence in Robotics and Automation, 2003. Proceedings. 2003 IEEE International Symposium on* (IEEE, 2003), vol. 2, pp. 954–959.

16. T. Ahonen, A. Hadid, M. Pietikäinen, *Computer vision-eccv 2004* pp. 469–481 (2004).

17. R. Adolphs, *Behavioral and cognitive neuroscience reviews* **1**, 21 (2002).

18. Y. Yacoob, L. S. Davis, *IEEE Transactions on pattern analysis and machine intelligence* **18**, 636 (1996).

19. G. Guo, L. Wen, S. Yan, *IEEE Transactions on Circuits and Systems for Video Technology* **24**, 814 (2014).

20. Z. Jing, R. Mariani, *Pattern Recognition, 2000. Proceedings. 15th International Conference on* (IEEE, 2000), vol. 2, pp. 933–936.

21. B. Wu, H. Ai, R. Liu, *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on* (IEEE, 2004), vol. 1, pp. 292–295.

22. J. A. M. Basilio, G. A. Torres, G. S. Pérez, L. K. T. Medina, H. M. P. Meana, *Applications of Mathematics and Computer Engineering* pp. 123–128 (2011).

23. C. Garcia, G. Tziritas, *IEEE Transactions on multimedia* **1**, 264 (1999).

24. K. Schwerdt, J. L. Crowley, *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on* (IEEE, 2000), pp. 90–95.

25. B. Schiele, A. Waibel, *Proceedings of the International Workshop on Automatic Face-and Gesture-Recognition* (1995), pp. 344–349.

26. A. Albiol, L. Torres, E. J. Delp, *Image Processing, 2001. Proceedings. 2001 International Conference on* (IEEE, 2001), vol. 1, pp. 122–124.

27. V. Vezhnevets, V. Sazonov, A. Andreeva, *Proc. Graphicon* (Moscow, Russia, 2003), vol. 3, pp. 85–92.

28. D.-H. Liu, K.-M. Lam, L.-S. Shen, *Pattern Recognition* **38**, 1705 (2005).

29. W.-C. Kao, M.-C. Hsu, Y.-Y. Yang, *Pattern Recognition* **43**, 1736 (2010).

30. K.-N. Kim, R. Ramakrishna, *Systems, Man, and Cybernetics, 1999. IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on* (IEEE, 1999), vol. 2, pp. 324–329.

31. Y. Yacoob, L. S. Davis, *IEEE transactions on pattern analysis and machine intelligence* **28**, 1164 (2006).

32. F. Volken, J. Terrier, P. Vandewalle, *Conference on Colour in Graphics, Imaging, and Vision* (Society for Imaging Science and Technology, 2006), vol. 2006, pp. 359–364.

33. N. Tsumura, *et al.*, *ACM Transactions on Graphics (TOG)* **22**, 770 (2003).

34. O. Déniz, G. Bueno, J. Salido, F. De la Torre, *Pattern Recognition Letters* **32**, 1598 (2011).

35. R.-L. Hsu, M. Abdel-Mottaleb, A. K. Jain, *IEEE transactions on pattern analysis and machine intelligence* **24**, 696 (2002).

36. C. Chen, A. Dantcheva, A. Ross, *Biometrics (ICB), 2013 International Conference on* (IEEE, 2013), pp. 1–8.

37. A. Albiol, D. Monzo, A. Martin, J. Sastre, A. Albiol, *Pattern Recognition Letters* **29**, 1537 (2008).

38. X. Wang, T. X. Han, S. Yan, *Computer Vision, 2009 IEEE 12th International Conference on* (IEEE, 2009), pp. 32–39.

39. P. Yang, S. Shan, W. Gao, S. Z. Li, D. Zhang, *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on* (IEEE, 2004), pp. 356–361.

40. O. M. Parkhi, A. Vedaldi, A. Zisserman, *et al.*, *BMVC* (2015), vol. 1, p. 6.