Hair Segmentation for Color Estimation in Surveillance Systems

Aleš Křupka, Jiří Přinosil, Kamil Říha, Jiří Minář
Department of Telecommunications
Brno University of Technology
Brno, Czech Republic
e-mail: {akrupka,jiri.minar}@phd.feec.vutbr.cz
{prinosil,rihak}@feec.vutbr.cz

Malay Kishore Dutta
Amity School of Engineering and Technology
Amity University
Uttar Pradesh, India
e-mail: mkdutta@amity.edu

***Abstract* — The paper proposes a novel method for hair segmentation, that can be used in real-time video surveillance systems or multimedia services. The method utilizes an approach of video subtraction to obtain a person silhouette. Subsequently, the head part can be identified on the silhouette by exploiting information about face position. From the head, the skin is separated using floodfilling procedure and the hair area is determined as the difference between the head and the skin. The precision of the method is evaluated using manually extracted hair masks. The purpose of the segmentation method is to specify a hair area which can be then used for a hair color estimation. Therefore, the usability of the hair segmentation procedure is tested by a proposed scheme of hair color estimation.**

Keywords-segmentation, hair, color, face detection, video subtraction, floodfilling.

I. INTRODUCTION

In recent years, computer vision has become an inherent part of modern surveillance systems. Algorithms for pedestrian detection, people tracking or identity verification are common parts of these systems. This paper deals with analysis of hair in video sequences. Hair is classified to the category of soft biometric traits. This means that it cannot be used for a person identification by itself, but it can help the identification together with other soft biometric traits. Further, it can be used for division people into sub-groups, for example by assigning a hair color index to a person in a video-sequence. Then, such indices can be used as an additional filtering clue when searching in multimedia databases.

Firstly, we mention relevant works in this topic. In [1], hair area is detected based on sliding window which evaluates the hair of color. In [2], color and frequency information is used for creating seeds. Hair is then extracted using matting process using the seeds. In [3], hair is segmented using Graph-Cut and Loopy Belief Propagation. In [4], hair seeds are detected and a growing of hair region is applied based on color and texture features. In [5], the approach of seed identification and consecutive propagation is used. This procedure is done in two stages where the second stage uses a specific hair model based on the first stage results. In [6], the hair seed patches are obtained via active shape and active contours. These areas are then used to train a model of hair color and texture. According to this model, the final hair area is determined. In [7], selected hair and background seed regions are used for online support vector machines (SVM) model training. This model is then used to differentiate between other hair/background pixels. In [8], the coarse hair probability map is estimated and this map is consequently refined using Isomorphic Manifold Inference Method to get optimal hair region. In [9], part-based model is proposed together with a way of modeling relations between the parts of head and hair which helps to a better hair identification.

The previous works are designated to work with static images and therefore the need to distinguish between a head and a background exists. The motivation of this work is to be able to estimate a hair color of people in a video-sequence, so this soft biometric trait can be extracted in real-time. The fact of using video sequence significantly simplifies the solution of head/background separation and therefore the hair segmentation procedure can be simplified in advance of shortening the processing time.

The paper is organized as follows: Section II describes the proposed method for hair region selection. Section III presents the experimental setup and the results of hair region selection and its usability for hair color estimation. Section IV then concludes the results and according to them proposes a direction of the future work.

II. HAIR SEGMENTATION METHOD

The method presumes the usage of video-sequences. The scheme of the method can be seen in Fig. 1. A hair is determined as a difference between head and skin area of a head. Every frame of the video-sequence is examined by a face detector for a face occurrence and it is also supplied to a background subtractor. If a face in the frame is detected, the position of the face is used for a segmentation of head. A silhouette of the person is obtained from the background subtractor and the head is given as the part of the silhouette. This part is specified by the position returned by the face detector. As the head mask is given, the skin area in the head is needed to be found. This is performed utilizing information about eyes position and nose position. This information is obtained during the face detection stage and it is used for a selection of proper points, which are used as seeds for a flooding procedure. Using the flooding procedure, the skin area is defined. Finally, the hair mask is given as the difference between the head segment and the skin segment.

Copyright (c) IARIA, 2014.

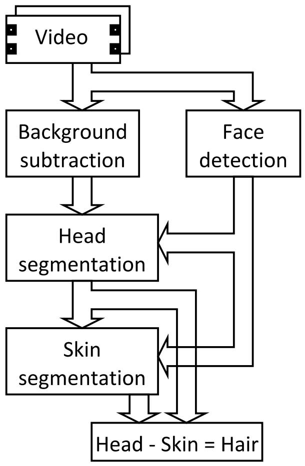

Figure 1. Scheme of the proposed method

A. Face detection

For the face detection, the widely used cascade Viola-Jones detector [10], implemented in OpenCV library, is used. Using the detector, a face occurrence in a frame can be located and a sub-window containing face is specified. Then, in the selected sub-window, eyes and nose are also located by the Viola-Jones detector, as it can be seen in Fig. 2(b). The position of eyes and nose is then used for further processing. In this stage, a successful extraction of face, eyes and nose is required for the further processing. Thus, if no face is detected in a frame or if eyes and nose positions are not obtained successfully, the processing procedure of the current frame is cancelled and the current frame is only supplied for the background subtractor. The processing procedure is then started with the following frame in the video-sequence.

B. Background modeling

Background modeling is a big advantage when using a video-sequence for the hair segmentation. The frames of the sequence allow to model a background scene and thus a silhouette of a moving human subject can be obtained. For this purpose, the method using Gaussian mixtures for background modeling [11] implemented in OpenCV library is used. This implementation also addresses shadow detection, thus shadows appearances can be eliminated during silhouette extraction. However, because the segmentation of a moving object in a frame using this method is still not perfect, the morphological opening is applied to remove small spurious segments in the frame. The moving object is then separated from the scene, as shown in Fig. 2(c).

C. Head segmentation

A head in the frame is obtained using the silhouette from the background subtractor and the face position from the face detector. The face position is presented by a rectangle with face. This rectangle is enlarged in order to cover the whole head area, as illustrated in Fig. 2(b). The size of the rectangle is enlarged by factor 1.5 which was empirically selected as optimal value. This rectangle thus contains a part of the silhouette corresponding to the head. Usually, the silhouette obtained from the background subtractor is not ideal. Concretely, it can consist of unconnected regions and thus the part corresponding to the head cannot be used as a head mask directly, as can be seen Fig. 2(c). Thus, a convex hull is constructed from the point set which is given as the union of regions in the head area. This convex hull then represents the head mask. The convex hull is constructed only from the pixels of the upper rectangle part to avoid including pixels of arms, the resulting head mask is shown in Fig. 3(a). Although the head mask does not cover all the head area, for the purpose of color estimation it is sufficient to have hair from the top of the head.

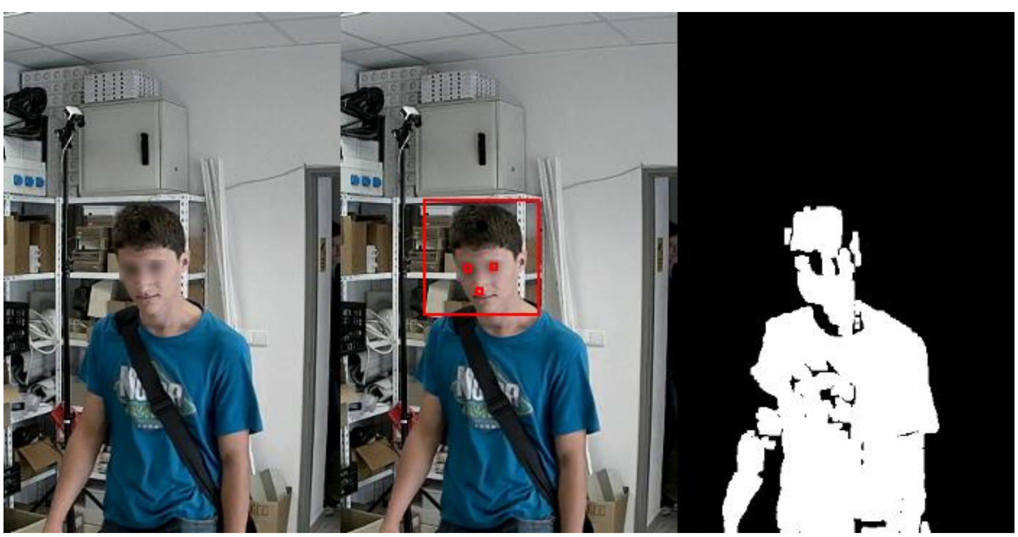

Figure 2. Illustration of initial video frame processing: (a) original frame, (b) face, eyes and nose detection, (c) silhouette extraction

D. Skin segmentation

The skin segmentation is the most crucial part of the processing. When a skin of a head is segmented correctly, then a difference between the head mask and the skin mask defines a hair mask. As mentioned in the beginning, the floodfilling approach is used for the skin segmentation. The flooding starts repeatedly from the different seed-points. For the floodfilling, an optimal RGB range was empirically selected. A pixel belongs to this range if the distance between its value and the value of the current seed-point is not greater than 30 (for the corresponding 8-bit color channels). The seed-points can be seen in Fig. 3(b), and they are given as points above and under eyes. Their positions are calculated utilizing the positions of the eyes and the nose. The reason for the usage of multiple seed-points is that the skin color varies in the different places of a face and thus the floodfilling would not work satisfactorily. An example of the color variation can be seen in Fig. 3 where one half of the face is darker due to shadows. When using multipleseed-points, the area with similar color around a particular seed-point is flooded. Then, the skin area is obtained as an union of the areas flooded from the different seed-points.

Similar as in the case of head segmentation, such the union of flooded areas does not give an ideal skin mask, as can be seen in Fig. 3(c). For example, the eyes' area is not flooded because the pixels' values are too different from seed point values. Thus, the same approach as during head segmentation, i.e., a convex hull of an union of flooded areas, is constructed. The final hair area is then given as the difference between the head and the skin area, as in Fig. 3(d).

Sometimes, the upper seed-points can fall into the hair area. In this case, such the seed-point cannot be used to initiate the floodfilling procedure. For a selection of proper seed-points, the color modeling method, described in [12], is used. The pixels of the line going through the lower seed-points are considered to have a color of skin, because the pixels are in the area around the nose. Thus, the color model of the skin is created using these pixels. If a color of a particular upper seed-point fits into this model, than the seed-point can be used to initiate the floodfilling procedure.

Finally, one more fact needs to be regarded. When a color of hair is similar to a skin color, the flood can also propagate into an area of hair, as can be seen in Fig. 4. Therefore, to obtain the correct skin segment, only a part of flooded area around the current seed-point is selected. The pixels around the seed-point are considered to be a skin as long as the shape of flood shrinks when going away from the seed point. This way, only a relatively compact part of the flooded area is selected, the selection is illustrated as blue areas in Fig. 4. This is based on the assumption that a texture of skin is more homogeneous than a texture of hair, so theflooding procedure forms more compact shapes on skin parts. Here follows the selection principle of a compact part: The flooded area is represented by the binary shape as shown in Fig. 5(a). Further, the distance transform of morphological type [13] is applied on this shape to get distance image illustrated in Fig. 5(b). In this image, the appropriate regional maximum is found according to the position of the current seed-point (black pixel in Fig. 5(b)). From this regional maximum, the process of descending to lower levels of distance image is performed in individual steps. At the beginning, the appropriate regional maximum is marked as positive. The other regional maxima are marked as negative. Then, the descending starts. In every step, pixels of a current level are marked as positive or negative. The pixels neighboring to negative pixels are marked as negative. The other pixels of the current level connected with positive pixels are marked as positive. These steps are repeated until the level of one is reached. The compact part is then composed from positive pixels as can be seen in Fig. 5(c).

Figure 3. (a) head segmentation, (b) seed-points, (c) union of flooded areas, (d) hair and skin segment

Figure 4. Selection of compact subpart of flood for different seed-points

III. EXPERIMENTS

The performance of the method was tested using 28 video sequences of average length of 4 seconds and 25 frames per second. Each video-sequence contains one person passing through a room, as can be seen in Fig. 2. On average, 37 hair masks are segmented from each video-sequence (a mask is obtained, if a face together with positions of eyes and nose is detected in a frame). These masks represent a hair of a person passing through the room. Totally, 1035 hair masks (extracted from all 28 video-sequences) were evaluated in this experiment.

The stages of face detection and background subtraction were performed on frames of size 854x480 pixels. To provide good stability of the background subtractor, several seconds of a video-sequence capturing the typical changes in the scene are supplied to the background subtractor to better model the background scene. When executing the part of skin segmentation, a square with detected face is normalized to size of 350x350 pixels. The part of selecting compact flooded parts around seed-points (Fig. 4) is due to computational reasons performed on down-sampled squares of 130x130 pixels. This step of down-sampling shortens the processing time.

A. Hair segmentation

As a reference, the area of hair was manually labeled.

Copyright (c) IARIA, 2014. ISBN: 978-1-61208-320-9

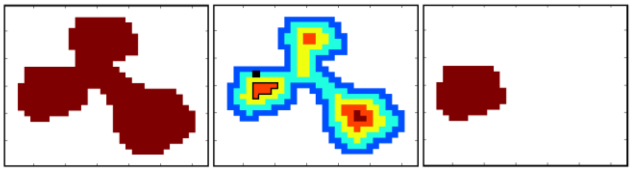

Figure 5. (a) binary shape, (b) distance image, (c) resulting selection

A mask extracted by using the described segmentation procedure is then compared with label mask and pixel counts such as true positives (*TP*), false positives (*FP*), false negatives (*FN*) and true negatives (*TN*) are given. Using these counts, the measures

$$precision = \frac{TP}{TP + FP} \quad (1)$$

and

$$recall = \frac{TP}{TP + FN} \quad (2)$$

can be computed.

Resulting hair masks often include on their border parts also non hair pixels. To alleviate this, the morphological erosion is applied on the resulting mask. Although this operation reduces recall measure, it increases precision measure. In the context of the utilization of segmented hair area for hair color estimation, precision can be considered the important measure, because a big number of false positive pixels (low precision) can more influence the final color estimation. On the other hand, the situation is not critical, when a subpart of hair is not included into the final color estimation (low recall), if we assume, that the hair area does not have more parts of different colors. The results of the experiments can be seen in Table. I.

TABLE I. SEGMENTATION RESULTS

Measure	***Average***	***Median***
Precision	0.84%	0.92%
Recall	0.49%	0.48%

The obtained results can be considered as good. Despite recall being quite low, it can be stated, that still approximately one half of hair area is marked by the segmentation procedure. On the other hand, the precision value is very good and it is comparable with the state of the art techniques referenced in the introduction. The main contribution is that the method was intended to be applicable in real-time systems and it is able to work sufficiently fast, see processing times in Table II.

TABLE II. PROCESSING TIMES

Phase	***Average time* [ms]**
Face detection	121
Background subtraction	55
Head and skin segmentation	38

The test was run on machine with Intel Core i5 M560 processor. Although the face detection stage is the most time consuming, the need of face position can be satisfied by frame down-sampling or face detection with rate lower than actual video frame rate. A face position in frames, where the detection is not done, are obtained using face tracking via optical flow as proposed in [14]. If we have a position of face and a silhouette by hand, the hair segmentation procedure is very fast. Therefore, the method can be implemented as an additional processing in various surveillance systems, where the face detection and background subtraction is also a part of processing and therefore, the hair segmentation method will not present a big additional computational burden.

B. Color estimation

In this stage, the suitability of the hair segmentation procedure for color estimation was tested. A hair color of 28 people in available video-sequences was estimated. Firstly, the color of hair was estimated from the area which is determined by hair segmentation procedure. Secondly, the color of hair was estimated from the area which is given by manually extracted hair label. To get a reference, a hair color of each person was subjectively classified. The colors, which are estimated from the pixels of extracted hair masks and hair labels, are then compared with this reference. The color can be classified into 5 different categories (black, brown, blond, red, gray/white).

For the automatic color classification, the following scheme is used. The RGB space can be considered as a cube. When it is equally spaced to 10 ranges for every dimension, 1000 sub-cubes are obtained. The color from the central RGB triplet of each sub-cube was subjectively evaluated using the previously mentioned color classes (plus a class, when a color is not a color of hair). Then, according to this evaluation, every sub-cube presents a range of RGB values corresponding to a given color - this approach assumes, that all possible colors in one sub-cube are similar.

A color of pixel, which is determined by the segmentation procedure as a hair pixel in a particular frame, is estimated using the scheme described above. When examining all hair pixels in the frame, a hair color histogram, showing numbers of pixels of defined hair colors, is obtained. For one person, every frame of the video-sequence is examined this way. The numbers of corresponding colors are then summed for all examined frames in the video-sequence. The biggest number determines the hair color of the person captured in the video-sequence. This procedure was conducted for all 28 videos. The estimation results, which are based on the hair masks extracted by the proposed hair segmentation method, are shown in Table III. For a comparison, Table IV provides results, when the color estimation is based on the pixels determined by manually extracted hair labels. It can be seen that the results are very similar, they differ only in two cases. Thus, we can state, that the hair segmentation method can be considered as good for purposes of hair color estimation, because the low value of recall only slightly influences the results of the color estimation procedure.

Copyright (c) IARIA, 2014. ISBN: 978-1-61208-320-9

TABLE. III COLOR ESTIMATION RESULTS USING EXTRACTED HAIR MASKS

		Estimation					
		Blk.	Brw.	Bld.	Red	Wht.	Perc.
Subj. Evaluation	Blk.	**13**	1	0	0	0	92.9%
	Brw.	2	**5**	0	0	1	62.5%
	Bld.	0	0	**0**	0	3	0%
	Red	1	0	0	**0**	0	0%
	Wht.	0	0	0	0	**2**	100%

TABLE. IV COLOR ESTIMATION RESULTS USING MANUALLY EXTRACTED HAIR LABELS

		Estimation					
		Blk.	Brw.	Bld.	Red	Wht.	Perc.
Subj. evaluation	Blk.	**13**	1	0	0	0	92.9%
	Brw.	1	**6**	0	0	1	75.0%
	Bld.	0	1	**0**	0	2	0%
	Red	1	0	0	**0**	0	0%
	Wht.	0	0	0	0	**2**	100%

For the discussion about usability of our hair color estimation model, we use Table IV. When we aim at discriminability of individual colors using our proposed model, we can see that the black and white/gray color estimation is the most successful (the people subjectively evaluated as having black or white/gray hair were correctly estimated in 92.9% or 100%, respectively). On the other hand, the subjectively perceived blond color was not correctly estimated in any case. According to the results, the blond color is estimated as white or brown. The reason is that subjectively evaluated blond hair can be in fact blend of more colors, the most common are white and brown. The brown color is correctly estimated in 75.0%. The red color is not correctly estimated in any case, however, only one person with subjectively evaluated red hair was present in the database, so the result of the red color cannot be considered as representative.

Although the color estimation scheme is rather simple, it was intended mainly for the evalution of usability of the developed hair segmentation method. However, as can be seen from the results, hair color estimation itself is a very challenging topic and therefore it will need further effort to develop sufficiently robust method, but this development is out of the scope of this paper. Generally, retrieval of color from image is not a trivial task. RGB values in image can be for example influenced by various sources of illuminations. Thus, an illuminant estimation and its inclusion into the color estimation procedure is needed as mentioned in survey [15]. For the future work, more sophisticated schemes for hair color estimation need to be developed such as the scheme presented in [16], where the color values are classified into predefined categories specified by google search engine.

IV. CONLUSION AND FUTURE WORK

The main objective of this paper was the proposal of the hair segmentation method which will be suitable for the following task of hair color estimation. This method analyses a hair of human from the frontal view and it is applicable on video-sequences. The main contribution of this proposal is its applicability in real-time systems because of its low computational requirements.

The method was tested on 28 labeled video-sequences and the results showed that precision, which is important measure in context of color estimation, has very good values. The relatively low recall of the proposed segmentation method has no big importance in the context of hair color estimation. This was supported by the experiment of the estimation of hair color. The resulting hair color was based on pixels which were determined by automatically extracted hair masks in the first case and by manually extracted hair labels in the second case. For both cases, the color estimation was nearly the same. For the estimation, the hair color model using equidistantly divided RGB space was applied. However, the hair segmentation method can be further developed especially to be able to recognize bald people. Currently, due to the segmentation errors, it is still not easy to distinguish bald people form the people with big forehead and thin layer of hair (bald people were not contained in video-sequences for evaluation).

From the results, some suggestions for the future work can be made. The color estimation scheme was rather simple and it was used to evaluate usability of the proposed hair segmentation procedure for the purpose of hair color analysis. As mentioned earlier, it is not a trivial task to automatically determine a color of hair as it is commonly done by humans. Therefore, a more sophisticated color estimation scheme should be proposed. This scheme will take into account a typical color composition of different hair color and also consider the influence of illumination. For example, some kind of supervised machine learning techniques combined with known methods of illuminant estimation could be utilized for these purposes. These techniques could take into account percentage composition of colors for a particular hair color class. Further, the reference hair colors were obtained by a subjective classification. Thus, because of the subjective classification, these values should be acquired from more subjects in standardized conditions to compare opinions of hair color from different evaluators. Eventually, the bigger experimental database should be acquired. This database should be balanced, when considering different hair colors, to better evaluate the obtained results.

ACKNOWLEDGMENT

This research was supported by part of the project reg. no CZ.1.07/2.3.00/20.0094 which is co-financed by the European Social Fund and the state budget of the Czech Republic and by the project VG20102014033 financed by the Ministry of the interior of the Czech Republic and by the project SIX CZ.1.05/2.1.00/03.0072.

References

[1] Y. Yacoob and L. S. Davis, "Detection and Analysis of Hair," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 7, July 2006, pp. 1164-1169.

[2] C. Rousset, "Frequential and color analysis for hair mask segmentation," 15th IEEE International Conference on Image Processing, Oct. 2008, pp. 2276 – 2279.

[3] K-C. Lee, D. Anguelov, B. Sumengen, and S. B. Gokturk, "Markov random field models for hair and face segmentation," 8th IEEE International Conference on Automatic Face & Gesture Recognition, Sept. 2008, pp. 1-6.

[4] U. Lipowezky, O. Mamo, and A. Cohen, "Using integrated color and texture features for automatic hair detection," IEEE 25th Convention of Electrical and Electronics Engineers in Israel, Dec. 2008, pp. 51-55.

[5] D. Wang, S. Shan, W. Zeng, H. Zhang, and X. Chen, "A novel two-tier Bayesian based method for hair segmentation," 16th IEEE International Conference on Image Processing, Nov. 2009, pp. 2401-2404.

[6] P. Julian, C. Dehais, F. Lauze, V. Charvillat, A. Bartoli, and A. Choukroun, "Automatic Hair Detection in the Wild," D. Wang, S. Shan, W. Zeng, H. Zhang, and X. Chen, "A novel two-tier Bayesian based method for hair segmentation," 20th International Conference on Pattern Recognition, Aug. 2010, pp. 4617-4620.

[7] D. Wang, X. Chai, H. Zhang, H. Chang, W. Zeng, and S. Shan, "A novel coarse-to-fine hair segmentation method," IEEE International Conference on Automatic Face & Gesture Recognition and Workshops, March 2011, pp. 233-238.

[8] D. Wang, S. Shan, H. Zhang, W. Zeng, and X. Chen, "Isomorphic Manifold Inference for hair segmentation," 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, April 2013, pp. 1-6.

[9] N. Wang, H. Ai, and F. Tang, "What are good parts for hair shape modeling?," IEEE Conference on Computer Vision and Pattern Recognition, June 2012, pp. 662 - 669.

[10] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Dec. 2001, pp. 511-518.

[11] Z. Zivkovic, "Improved adaptive Gausian mixture model for background subtraction," Proceedings of the 17th International Conference on Pattern Recognition, Aug. 2004, pp. 28-31.

[12] T. Horprasert, D. Harwood, and L. S. Davis, "A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection," International Conference on Computer Vision FRAME-RATE Workshop, 1999, pp. 1- 19.

[13] L.Vincent, "Granulometries, Segmentation, and Morphological Algorithms," Lecture Notes for Morphological Image and Signal Processing Workshop, September 1995, pp. 37-41.

[14] K. Říha, J. Přinosil, D. Fu, M. Zukal, and J. Karásek, "Method for Real-Time Face Tracking in a Video Sequence", Advances in Sensors, Signals, Visualisation, Imaging and Simulation. Recent Advances in Electrical Engineering Series, 2012, pp. 182-187.

[15] A. Gijsenij, T. Gevers, and J. van de Weijer, "Computational Color Constancy: Survey and Experiments," IEEE Transactions in Image Processing, vol. 20, no. 9, Sept. 2011, pp. 2475 - 2489.

[16] J. van de Weijer, C. Schmid, J. Verbeek, and D. Larlus, "Learning Color Names for Real-World Applications," IEEE Transactions in Image Processing, vol. 18, no. 7, July 2009, pp. 1512 - 1523.