A New Robust Color Descriptor for Face Detection

Eyal Braunstain and Isak Gath
Faculty of Biomedical Engineering, Technion - Israel Institute of Technology, Haifa, Israel

Keywords: Face detection, Object recognition, Descriptors, Color.

Abstract: Most state-of-the-art approaches to object and face detection rely on intensity information and ignore color information, as it usually exhibits variations due to illumination changes and shadows, and due to the lower spatial resolution in color channels than in the intensity image. We propose a new color descriptor, derived from a variant of Local Binary Patterns, designed to achieve invariance to monotonic changes in chroma. The descriptor is produced by histograms of encoded color texture similarity measures of small radially-distributed patches. As it is based on similarities of local patches, we expect the descriptor to exhibit a high degree of invariance to local appearance and pose changes. We demonstrate empirically by simulation the invariance of the descriptor to photometric variations, i.e. illumination changes and image noise, geometric variations, i.e. face pose and camera viewpoint, and discriminative power in a face detection setting. Lastly, we show that the contribution of the presented descriptor to face detection performance is significant and superior to several other color descriptors, which are in use for object detection. This color descriptor can be applied in color-based object detection and recognition tasks.

1 INTRODUCTION

Most object and face detection algorithms rely on intensity-based features and ignore color information. This is usually due to its tendency to exhibit variations due to illumination changes and shadows (Khan et al., 2012a), and also to the lower spatial resolution in color channels than in the intensity image (e.g. the works of (Viola and Jones, 2004; Mikolajczyk et al., 2004; Zhang et al., 2007; Li et al., 2013)). Face detection performance by a human observer declines when color information is removed from faces (Bindemann and Burton, 2009). It has been argued that a detector which is based solely on spatial information derived from an intensity image, e.g. histograms of gradients, may fail when the object exhibits changes in spatial structure, e.g. pose, non-rigid motions, occlusions etc. (Wei et al., 2007). Specifically, an image color histogram is rotation and scale-invariant.

We hereby review the topic of color representations and descriptors for object detection. Color information has been successfully used for object detection and recognition (Khan et al., 2012a; Gevers and Smeulders, 1997; Weijer and Schmid, 2006; Diplaros et al., 2006; Khan et al., 2013; Van de Sande et al., 2010; Wei et al., 2007; Khan et al., 2012b).

Color can be represented in various color spaces, e.g. RGB, HSV and CIE-Lab, in which uniform changes are perceived uniformly by a human observer (Jain, 1989). Various color descriptors can be designed. The color bins descriptor (Wei et al., 2007) is composed of multiple 1-D color histograms by projecting colors on a set of 1-D lines in RGB space at 13 different directions. These histograms are concatenated to form the color bins features.

Two color descriptors were examined by (Weijer and Schmid, 2006) for object detection, the Robust Hue descriptor, invariant with respect to the illuminant variations and lighting geometry variations (assuming white illumination), and Opponent Angle (OPP), invariant with respect to illuminant and diffuse lighting (i.e. light coming from all directions).
The trade-off between photometric invariance and discriminative power was examined in (Khan et al., 2013), where an information theoretic approach to color description for object recognition was proposed. The gains of photometric invariance are weighted against the loss in discriminative power. This is done by formulation of an optimization problem with objective function based on KL-Divergence between visual words and color clusters.

Deformable Part Model (DPM) is used to model objects using spring-like connections between object parts (Felzenszwalb et al., 2010; Zhu and Ramanan, 2012). Although DPM achieves very good detection results, in particular through its ability to handle

Braunstain E. and Gath I..
A New Robust Color Descriptor for Face Detection.
DOI: 10.5220/0005177400130021
In *Proceedings of the International Conference on Pattern Recognition Applications and Methods* (ICPRAM-2015), pages 13-21
ISBN: 978-989-758-077-2

challenging objects (e.g. deformations, view changes and partially occluded objects), the general computational complexity of part-based methods is higher than global feature-based methods (Bergtholdt et al., 2010; Heisele et al., 2003).

The Three-Patch Local Binary Patterns (TPLBP) (Wolf et al., 2008) is a robust variant of the Local Binary Patterns (LBP) descriptor (Ojala et al., 2002), based on histograms of encoded similarity measures of local intensity patches. This descriptor was examined for the face recognition task.

In the present work the focus is not on the design of a new face detection framework, but rather on the design of a novel color descriptor, investigating its possible contribution to face detection. We design a new color descriptor, based on Three-Patch LBP. Our descriptor is computed from histograms of encoded similarities of small local patches of chroma channels in a compact form, utilizing the inter-correlation between image chroma channels. Consequently, the representation of color in an image window is global, i.e. not part-based. We examine the descriptor by ways of its robustness to photometric and geometric variations and discriminative power. We evaluate the contribution of the descriptor in a face detection setting, using the FDDB dataset (Jain and Learned-Miller, 2010), and show that it exhibits significant contribution to detection rates.

The paper is organized as follows. In Section 2 the Three-Patch LBP (TPLBP) descriptor is described briefly, and a multi-scale variant is proposed; in section 3 the new color descriptor is described; in section 4 invariance and discriminative power are evaluated, compared to the Robust Hue and Opponent Angle descriptors (Weijer and Schmid, 2006); in section 5 we evaluate the color descriptor in a face detection setting, and in section 6 conclusions to this work are provided.

2 THREE-PATCH LBP DESCRIPTOR AND A MULTI SCALE VARIANT

The Three-Patch LBP (Wolf et al., 2008) descriptor was inspired by the Self-Similarity descriptor (Shechtman and Irani, 2007), which compares a central intensity image patch to surrounding patches from a predefined area, and is invariant to local appearance. For each central pixel, a $w \times w$ patch is considered, centered at that pixel, and S additional patches distributed uniformly in a ring of radius r around that pixel. Given a parameter α (where $\alpha < S$), we take S pairs of patches, α-patches apart, and compare their values to the central patch. A single bit value for the code of the pixel is determined according to which of the two patches is more similar to the central patch. The code has S bits per pixel, and is computed for pixel p by:

$$TPLBP\,(p) = \textstyle\sum_{i=1}^{S} f_\tau \left(d\left(C_i, C_p\right) - d\left(C_{i'}, C_p\right)\right) \cdot 2^i$$

$$i' = (i+\alpha)\ mod\,S \tag{1}$$

where C_i and $C_{(i+\alpha)\,mod\,S}$ are two $w \times w$ patches along the patches-ring, α-patches apart, C_p is the central patch, $d\left(\cdot,\cdot\right)$ is a distance measure (metric), e.g. L_2 norm, and the function f_τ is a step threshold function, $f_\tau(x) = 1$ iff $x \geq \tau$. The threshold value τ is chosen slightly larger than zero, to provide stability in uniform regions. The values in the TPLBP code image are in the range $\left[0, 2^S - 1\right]$. Different code words designate different patterns of similarity. Once the image is TPLBP-encoded, the code image is divided into non-overlapping cells, i.e. distinct regions, and a histogram of code words with 2^S bins is constructed for each cell. The histograms of all cells are normalized to unit norm and concatenated to a single vector, which constitutes the TPLBP descriptor.

We propose a Multi-Scale TPLBP descriptor (termed TPLBP-MS), capturing spatial similarities at various scales and resolutions, by concatenating TPLBP descriptors with various parameters r and w. The scale is affected by the radius r and patch resolution by patch size w. Three sets of parameters are used for the encoding operator of Eq. (1), i.e. $(r,S,w) = \{(2,8,3),(3,8,4),(5,8,5)\}$, all with $S = 8$ and $\alpha = 2$, as in (Wolf et al., 2008). These 3 TPLBP descriptors are concatenated to produce the TPLBP-MS descriptor. Parameters r and w are changed in similar manner in the 3 sets above, thus observing larger scales at lower resolutions.

3 A NEW COLOR DESCRIPTOR - COUPLED-CHROMA TPLBP

Many color descriptors are histograms of color values in some color space, e.g. rg-histogram and Opponent Colors histograms (Van de Sande et al., 2010). Image color channels contain texture information that is disregarded by color histograms. Our motivation is to formulate a color descriptor that captures the texture information embedded in color channels in a robust manner.

Color descriptors can be evaluated by several main properties: (1) Invariance to photometric changes (e.g. illumination, shadows etc.); (2) Invariance to geometric changes (e.g. camera viewpoint, object pose, scale etc.); (3) Discriminative power, i.e. the ability to distinguish a target object from the rest of the world; (4) Stability, in a sense that the variance of a certain dissimilarity measure between descriptor vectors of samples from a specific distribution (or class) is low. We would like to formulate a color descriptor that adheres to these properties.

We represent color in CIE-Lab space, due to its perceptual uniformity to a human observer. Using Euclidean distance in CIE-Lab space approximates the perceived distance by an observer, hence a detector based on this color space can in some sense approximate the perception of human color vision. In CIE-Lab space, L is the luminance, a and b are the chroma channels. We consider first a color descriptor produced by applying TPLBP to both chroma channels and concatenating the single-channel descriptors to a single descriptor. Images in JPEG format are analyzed, in which the chroma channels are sub-sampled (Guo and Meng, 2006), thus spatial resolution in chroma channels is lower than in intensity. Hence, to extract meaningful features from chroma, the appropriate operator should be applied at a coarse resolution, relative to the operator applied to the intensity image. The values of the parameters are chosen accordingly, $(r,S,w) = (5,8,4)$, i.e. both the radius and patch dimension are increased. This descriptor is termed Chroma TPLBP (C-TPLBP). It has twice the size of TPLBP.

A degree of correlation exists between the chroma channels in CIE-Lab space. This can be observed either from the derived equations of CIE-Lab color space from CIE-XYZ space, or from an experimental perspective, by constructing a 2-D chroma histogram of face images. Elliptically cropped face images from the FDDB dataset (Jain and Learned-Miller, 2010) with 2500 images are used to fit a 2-D Gaussian density of chroma values a and b by mean and covariance of the data. From the covariance matrix, we have that $\sigma_{ab} = 53.7$, i.e. nonzero correlation between the chroma channels. We presume that coupling the chroma channels information may lead to a robust descriptor, which is also more compact than C-TPLBP, where chroma channels descriptions are computed separately. We propose the following operator:

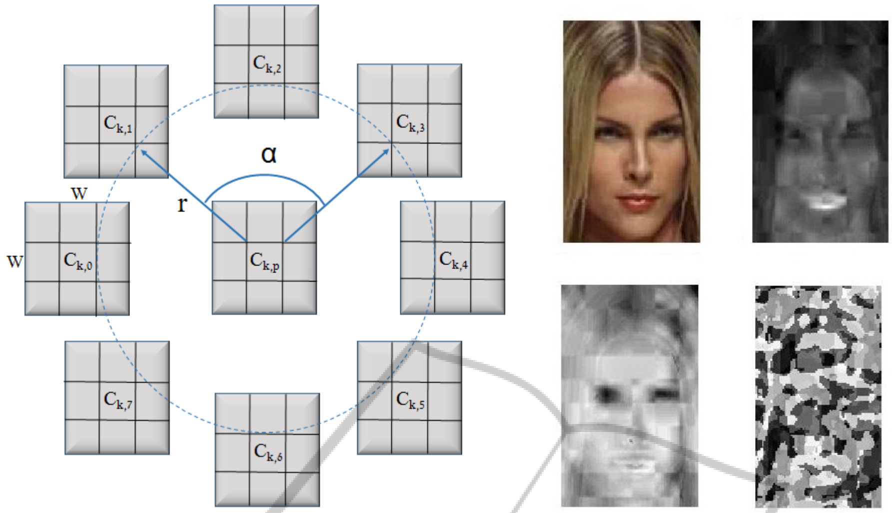
$$CC-TPLBP(p) = \sum_{i=1}^{S} f_\tau \left(\sum_{k=a,b} \left(d\left(C_{k,i}, C_{k,p}\right) - d\left(C_{k,i'}, C_{k,p}\right) \right) \right) \cdot 2^i$$

$$i' = (i+\alpha)\, mod\, S \tag{2}$$

where $C_{k,i}$ is the ith patch of chroma channel k and the inner summation is over chroma channels, a and b. The thresholding function f_τ operates on the sum of differences of patches distance functions, for both chroma channels. Given a parameter α, we take S pairs of patches from each chroma channel, α-patches apart, and for each pair we compare distances to the central patch of the appropriate channel. A single bit value for the code of a pixel is determined as follows - if similarities in both chroma channels correlate, e.g. if in both chroma channels patch C_i is more similar to the central patch C_p than patch $C_{i+\alpha}$, then the appropriate bit will be assigned value 0 (value 1 in the opposite case). Conversely, if dissimilarities of the two channels do not correlate, then by viewing the argument of the function f_τ as $\sum_{k=a,b} d\left(C_{k,i}, C_{k,p}\right) - \sum_{k=a,b} d\left(C_{k,(i+\alpha)\, mod\, S}, C_{k,p}\right)$, the patch with lower sum of distances in both chroma channels is more similar to the center, and the code bit is derived accordingly. The computed code has S bits per pixel, and this descriptor is of the same size as TPLBP, i.e. half the size of C-TPLBP. This descriptor is termed Coupled-Chroma TPLBP (CC-TPLBP). The parameters are chosen in accordance with those of C-TPLBP, $(r,S,w) = (5,8,4)$ and $\alpha = 2$. We emphasize that different values for the radius (r), number of patches (S), patch dimension (w) and α may be chosen, however, preliminary experiments showed that good discriminative ability was obtained with the parameter values specified above. The histograms are computed on small cells of (20, 20) pixels, thus maintaining the spatial binding of color and shape information in the image by cells delimitation, i.e. late fusion of color and shape (Snoek, 2005; Khan et al., 2012a). CC-TPLBP is invariant to monotonic variations of chroma and luminance. Such variations do not cause any change to the resulting descriptor. In Fig. 1 we present the CC-TPLBP operator, where the index $k = \{a,b\}$ designates the chroma channel, as in Eq. (2), with an example code computation for a color face image. CC-TPLBP can be combined with intensity-based shape features for classification tasks.

4 EVALUATION OF COLOR DESCRIPTORS

CC-TPLBP is invariant to monotonic changes of both luminance and color channels. Moreover, we expect it to exhibit a high degree of robustness to geometrical changes, e.g. pose, local appearance and camera viewpoint, as it is computed by similarities of radially-distributed image patches. We evaluate CC-TPLBP with respect to properties (1) - (4) described in section 3, compared to the Robust Hue and Opponent Angle (OPP) color descriptors (Weijer and



(a) (b)

Figure 1: CC-TPLBP code computation. (a) CC-TPLBP operator for a single chroma channel, with parameters $\alpha = 2$, S=8 and w=3. (b) An example of CC-TPLBP code for a color face image. Upper left - face image; upper right - CIE-Lab *a* chroma; lower left - CIE-Lab *b* chroma (*a* and *b* are presented as gray-level images); lower right - CC-TPLBP code image. The parameters used are r=5, S=8, w=4

Schmid, 2006). Opponent Colors are invariant with respect to lighting geometry variations, and are computed from RGB by:

$$O1 = \frac{1}{\sqrt{2}}(R-G)$$
$$O2 = \frac{1}{\sqrt{6}}(R+G-2B) \tag{3}$$

The Robust Hue descriptor is computed as histograms on image patches over hue, which is computed from the corresponding RGB values of each pixel, according to:

$$hue = \arctan\left(\frac{O1}{O2}\right) = \arctan\left(\frac{\sqrt{3}(R-G)}{R+G-2B}\right) \tag{4}$$

Hue is invariant with respect to lighting geometry variations when assuming white illumination. Hue is weighted by the saturation, to reduce error. The Opponent Derivative Angle descriptor (OPP) is computed on image patches, by the histogram over the opponent angle:

$$ang_x^O = \arctan\left(\frac{O1_x}{O2_x}\right) \tag{5}$$

where $O1_x$ and $O2_x$ are spatial derivatives of the chromatic opponent channels. OPP is weighted by the chromatic derivative strength, i.e. by $\sqrt{O1_x^2 + O2_x^2}$, and is invariant with respect to diffuse lighting and spatial sharpness. Color histograms are generally considered more invariant to pose and viewpoint changes than shape descriptors (Diplaros et al., 2006), but are sensitive to changes of illumination and shading.

We evaluate invariance and discriminative power by the Kullback-Leibler Divergence, a non-symmetric dissimilarity measure between two probability distributions, p and q, expressed as:

$$D_{KL}(p\|q) = \sum_i p_i \log\left(\frac{p_i}{q_i}\right) \tag{6}$$

where q is considered a model distribution.

We consider descriptors that are constructed from M histograms of M distinct image cells. Referring to CC-TPLBP, each histogram has 2^S bins, producing a descriptor of size $M \times 2^S$. Given two images, each with M cells, we compute M histograms for each image. To compare CC-TPLBP descriptors of these two images, we compute the KL Divergence for each pair of appropriate histograms from both images, i.e. $\{D_{KL}(h_{1,m}, h_{2,m})\}_{m=1,..,M}$, where $\{h_{i,m}\}_{i=1,2}^{m=1,..,M}$ is the mth histogram of image i. We define the KL Divergence of image 1 with respect to image 2 by averaging over all image cells, i.e. $D_{KL}^{1,2} = \frac{1}{M}\sum_{m=1}^{M}(D_{KL}(h_{1,m}, h_{2,m}))$. Each single-cell histogram contains $2^8 = 256$ bins.

We evaluate the CC-TPLBP, Hue and OPP descriptors by three experiments, described as follows:

Figure 2: Example of face and background images used to examine discriminative power. Top line - sample face images, bottom line - sample background images.

4.1 Invariance to Photometric and Geometric Variations

In the first experiment we evaluate invariance to combined photometric and geometric variations, i.e. illumination and background, face pose and viewpoint. While this does not allow for independent evaluations of invariance to photometric and geometric variations, it simulates a realistic setting for face detection. We use several groups of images of single persons from the LFW Face Recognition dataset (Huang et al., 2007), each group displays a single person with the above variations. We compute the CC-TPLBP, Hue and OPP histograms for all images in a set, normalized to unit sum, and the KL Divergence between histograms of all image pairs (which is non-symmetric, i.e. $D_{KL}(p_i,p_j) \neq D_{KL}(p_j,p_i)$). Table 1 contains statistics of KL Divergence values of all descriptors for several image sets. While the number of images is relatively small, the number of resulting pairing is large and therefore indicative. CC-TPLBP appears to be most robust to these variations, as its mean KL Divergence is by far the lowest from all descriptors on all image sets. CC-TPLBP also exhibits a higher degree of stability than other descriptors, by its lowest variance.

4.2 Invariance to Gaussian Noise

In the second experiment, we test the effects of added noise, using 2500 face images from the FDDB dataset (Jain and Learned-Miller, 2010), normalized to size 63×39 pixels. According to (Diplaros et al., 2006), sensor noise is normally distributed, as additive Gaussian noise is widely used to model thermal noise, and is a limiting behavior of photon counting noise. High Gaussian noise is added to R, G and B channels of all images, i.e. $\{R,G,B\}_D = \{R+n^R_{xy}, G+n^G_{xy}, B+n^B_{xy}\}$, where $\{n^k_{xy} = n^k(x,y)\}_{k=R,G,B}$, $n(x,y) \sim \mathcal{N}(0,\sigma_n)$, with $\sigma_n = 5$. We calculate KL Divergence between descriptor histograms of original and corrupted images. Statistics of the KL Divergence values are displayed in Table 2. While Hue has an average KL Divergence slightly lower than CC-TPLBP, the latter has significantly lower variance than other descriptors, indicating higher stability under addition of Gaussian noise.

4.3 Discriminative Power

In the third experiment, we examine discriminative power. A descriptor based on color histograms would be effective in distinguishing face patches from distinct objects, e.g. trees or sky patches, but may be less effective in distinguishing a face from skin, e.g. neck, torso. Here a color texture descriptor may be more efficient. We choose randomly 200 face images from the FDDB dataset, and pick 200 background images (see supplementary material) that give a degree of diversity and challenge for the considered descriptors, i.e. versatility of chroma and texture. Half of the background images do not contain skin at all, and the other half partially contain skin, with variable backgrounds. This image set is constructed to represent the kind of natural setting where the function of the descriptor is to be able to discriminate face patches from non-face skin patches together with versatile non-skin background. Several examples are presented in Fig. 2.

To evaluate discriminative power, we use the KL Divergence similar to (Khan et al., 2012a). We define a KL-ratio for face sample, considering all face and background samples in the set:

Table 1: Statistics of KL-Divergence, combined evaluation of photometric and geometric invariance, for several sets of single-person images. KL-Divergence is calculated for all pairs of images in a set. For further explanation, see text.

Person set (No. images / No. pairs)	Descriptor	Mean	Median	STD
Jennifer Aniston (21 / 420)	CC-TPLBP	0.1034	0.096	0.0353
	Hue	0.9666	0.8321	0.5885
	OPP	0.3555	0.2637	0.288
Arnold Schwarzenegger (42 / 1722)	CC-TPLBP	0.1154	0.1	0.0519
	Hue	1.3682	1.2901	0.6468
	OPP	0.6535	0.482	0.551
Vladimir Putin (49 / 2352)	CC-TPLBP	0.1124	0.0988	0.0515
	Hue	1.1799	1.0515	0.6268
	OPP	0.467	0.3582	0.3737

Table 2: Statistics of KL-Divergence, noisy images.

Desc.	Mean	Median	STD
CC-TPLBP	0.0553	0.0543	0.0168
Hue	0.0492	0.0397	0.0405
OPP	0.0968	0.0818	0.0579

Table 3: Statistics of KL-ratios; discriminative power. CC-TPLBP is found most discriminative.

Desc.	Mean	Median	STD
CC-TPLBP	1.7402	1.7506	0.1678
Hue	1.456	1.3835	0.361
OPP	1.6671	1.6712	0.2185

$$KL-ratio_k = \frac{\frac{1}{N_B}\sum_{j\in B} KL(p_j, p_k)}{\frac{1}{N_F-1}\sum_{i\in F, i\neq k} KL(p_i, p_k)} \quad \forall k \in F \tag{7}$$

where p_k is the descriptor of face patch $k \in F$, p_j is the descriptor of background patch $j \in B$, N_F and N_B are the number of face and background samples, respectively. For a face sample k, Eq. (7) defines the ratio of the average KL Divergence with all non-face patches, divided by the average KL Divergence with all face patches. The higher this ratio for a face patch $k \in F$, the more discriminative the descriptor with respect to this face and data set, as the intra-class KL Divergence is lower than the inter-class KL Divergence. The KL-ratio values of all descriptors on the dataset are displayed in Fig. 3, after low-pass filtering by a uniform averaging filter of size 7. Smoothing is performed in order to reduce the noisiness in the original KL-ratio curves. Statistics of the KL-ratios (prior to low-pass filtering) are given in Table 3. We observe that the average KL-ratio for CC-TPLBP is higher than that of Hue and OPP (i.e. higher discriminative power), and that the variance of CC-TPLBP is the lowest, indicating high stability (i.e. low variability of KL-ratios for data samples from a specific class in a dataset).

5 EVALUATION OF THE COLOR DESCRIPTOR IN A FACE DETECTION SETTING

We evaluate the CC-TPLBP color descriptor in a face detection setting.

5.1 Dataset

We use the FDDB benchmark (Jain and Learned-Miller, 2010), which contains annotations of 5171 faces in 2845 images, divided into 10 folds. five folds are used for training, and five for testing. Training face images are normalized to size 63×39. The background set is constructed from random 63×39 - sized patches from background images of the NICTA dataset (Overett et al., 2008), i.e. of same size as the face patches.

5.2 Evaluation Protocol

In our face detection system, we use Support Vector Machines (Cortes and Vapnik, 1995), a classification method that has been successfully applied for face detection (Romdhani et al., 2004; Osuna et al., 1997), as the face classifier. We examine various descriptors combinations, i.e. (1) TPLBP, (2) TPLBP-MS, (3) TPLBP-MS + Hue, (4) TPLBP-MS + OPP, (5) TPLBP-MS + C-TPLBP and (6) TPLBP-MS + CC-TPLBP. For each of (1)-(6) we train a linear-kernel SVM classifier with Soft Margin, where the regularization parameter C is determined by K-fold cross-validation (K=5). To reduce false alarm rate, we add a confidence measure for an SVM classifier decision, as a probability for a single decision (Platt, 1999):

$$p(\mathbf{w}, \mathbf{x}, y) = \frac{1}{1+\exp(-y(\mathbf{w}\cdot\mathbf{x}+b))} \tag{8}$$

where $\mathbf{w}$ is the SVM separating hyperplane normal vector, $\mathbf{x}$ is a test sample and y is the classifi-

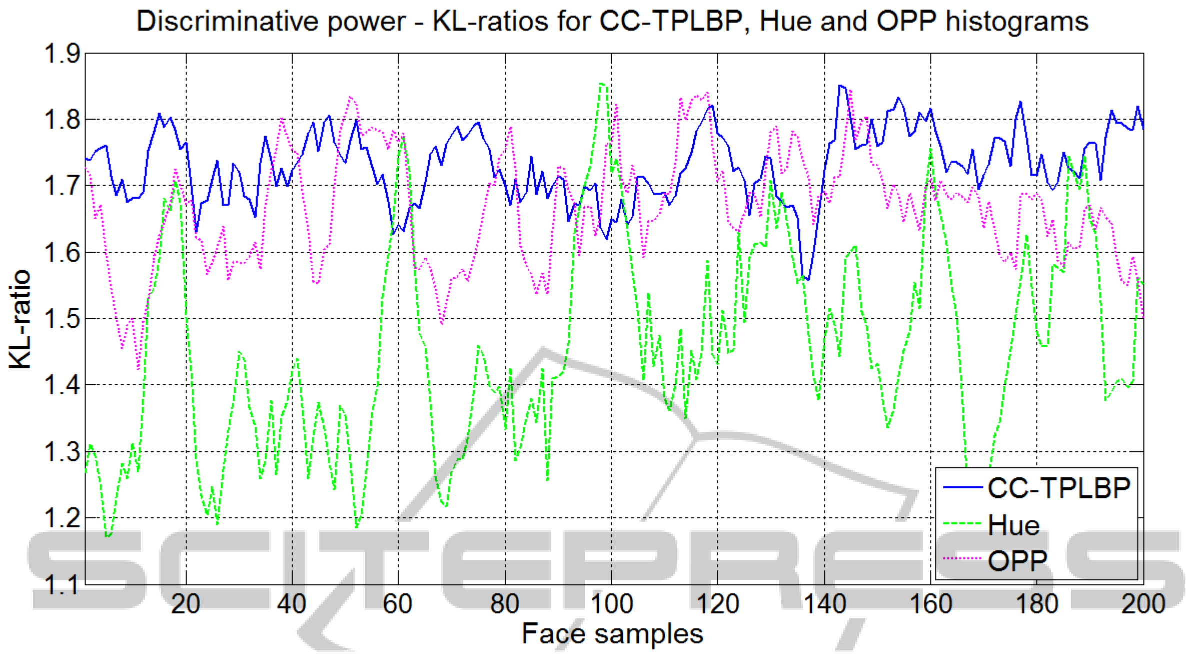


Figure 3: Discriminative power measure. KL-ratios of 200 face images with 200 background images. Horizontal axis: face sample numbers; vertical axis: KL-ratio values computed by Eq. (7). The displayed KL-ratios are smoothed using a uniform averaging filter of size 7, for further explanation see text. It can be seen that CC-TPLBP (blue curve) has the highest mean KL-ratio and lowest variance, as also seen in Table 3.

cation label. This logistic (sigmoid) function assigns high confidence (i.e. close to 1) to correctly-classified samples which are distant from the hyperplane.

Preprocessing of an image is performed by applying skin detection in CIE-Lab color space, to reduce image area to be scanned by a sliding window method. Various skin detection methods and color spaces can be used (hsuan Yang and Ahuja, 1999; Jones and Rehg, 2002; Zarit et al., 1999; Terrillon et al., 2000; Braunstain and Gath, 2013). We train offline a skin histogram based on chroma (a, b), omitting the luminance L as it is highly dependent on lighting conditions (Cai and Goshtasby, 1999). Skin detection in a test image is performed pixel-wise, by the application of threshold τ_s, i.e. for pixel $p = (x_p, y_p)$ with quantized chroma values $(\bar{a}_p, \bar{b}_p)$ and histogram value $h(\bar{a}_p, \bar{b}_p) = h_p$, the pixel is classified as skin if $h_p > \tau_s$. After skin is extracted, we perform a sliding window scan to examine windows at various positions and scales. The confidence measure of Eq. (8) is used by applying a threshold, i.e. if $p(\mathbf{w}, \mathbf{x}, y) > p_{th}$, the window is classified as a face.

5.3 Results

Face detection performance was evaluated by following the evaluation scheme proposed in (Jain and Learned-Miller, 2010). Receiver Operating Characteristic (ROC) were computed, with True Positive rate ($TPR \in [0,1]$) vs. number of False Positives (FP). In Fig. 4, ROC curves of continuous score (Jain and Learned-Miller, 2010) are depicted for various descriptor combinations. We observe that each of the descriptor combinations, TPLBP-MS, C-TPLBP and CC-TPLBP produce significant improvements in detection rates, compared to TPLBP. CC-TPLBP leads to similar performance as C-TPLBP, but with a more compact representation.

6 CONCLUSIONS

In the present work the focus is not on the design or optimization of a face detection framework, but rather on color representation, or description, for the task of face detection. We proposed a novel color descriptor, CC-TPLBP, which captured the texture information embedded in color channels. CC-TPLBP is by definition invariant to monotonic changes in chroma and luminance channels. A multi-scale variant of TPLBP is designed, termed TPLBP-MS. All experiments were performed in a face detection setting. We examined the invariance of CC-TPLBP, jointly for photometric and geometric variations, i.e. illumination, back-

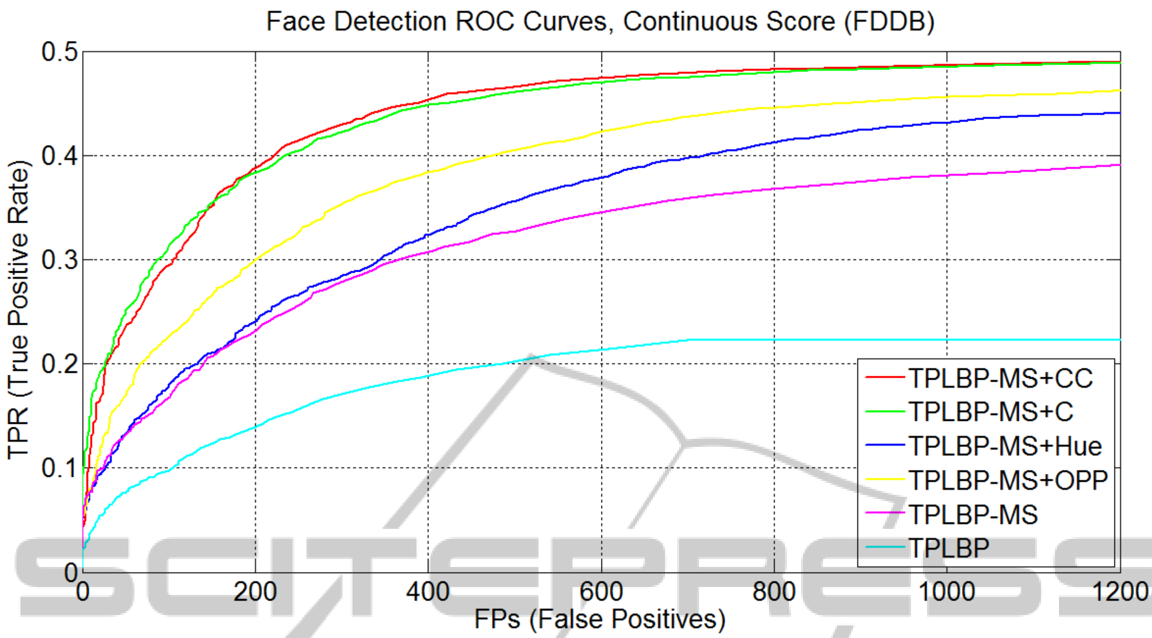


Figure 4: Face detection ROC curves on FDDB, for various descriptors combinations. It is clearly discerned that both CC-TPLBP and C-TPLBP (red and green lines, respectively) outperform all the other descriptor combinations. In addition, CC-TPLBP is twice more compact than C-TPLBP, making it the more efficient representation.

ground, face pose and viewpoint changes, and separately for addition of Gaussian noise, and compared to the Robust Hue and Opponent Angle (OPP) descriptors. Discriminative power was evaluated with respect to the above mentioned descriptors. CC-TPLBP is superior to the other two descriptors. It achieves higher discriminative power and much higher invariance to combined photometric and geometric variations, compared to Hue and OPP, as demonstrated in section 4. The evaluation experiments in a face detection setting demonstrated that (1) TPLBP-MS improves detection rates compared to TPLBP, (2) the addition of CC-TPLBP produces a sharp improvement over TPLBP-MS and (3) CC-TPLBP leads to superior detection rates compared to Hue and OPP.

The CC-TPLBP color-based descriptor can be integrated into face detection frameworks to achieve a substantial improvement in performance using existent color channels information. It can also be used in general color-based object recognition tasks.

REFERENCES

Bergtholdt, M., Kappes, J., Schmidt, S., and Schnörr, C. (2010). A study of parts-based object class detection using complete graphs. *Int. J. Comput. Vision*, 87(1-2):93–117.

Bindemann, M. and Burton, A. M. (2009). The role of color in human face detection. *Cognitive Science*, 33(6):1144–1156.

Braunstein, E. and Gath, I. (2013). Combined supervised / unsupervised algorithm for skin detection: A preliminary phase for face detection. In *Image Analysis and Processing - ICIAP 2013 - 17th International Conference, Naples, Italy, September 9-13, 2013. Proceedings, Part I*, pages 351–360.

Cai, J. and Goshtasby, A. A. (1999). Detecting human faces in color images. *Image Vision Comput.*, 18(1):63–75.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3):273–297.

Diplaros, A., Gevers, T., and Patras, I. (2006). Combining color and shape information for illumination-viewpoint invariant object recognition. *IEEE Transactions on Image Processing*, 15:1–11.

Felzenszwalb, P. F., Girshick, R. B., McAllester, D., and Ramanan, D. (2010). Object detection with discriminatively trained part-based models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32:1627–1645.

Gevers, T. and Smeulders, A. (1997). Color based object recognition. *Pattern Recognition*, 32:453–464.

Guo, L. and Meng, Y. (2006). Psnr-based optimization of jpeg baseline compression on color images. In *ICIP*, pages 1145–1148. IEEE.

Heisele, B., Ho, P., Wu, J., and Poggio, T. (2003). Face recognition: Component-based versus global approaches.

hsuan Yang, M. and Ahuja, N. (1999). Gaussian mixture model for human skin color and its applications in image and video databases. In *Its Application in Image* *and Video Databases. Proceedings of SPIE 99 (San Jose CA*, pages 458–466.

Huang, G. B., Ramesh, M., Berg, T., and Learned-Miller, E. (2007). Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst.

Jain, A. K. (1989). *Fundamentals of Digital Image Processing*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Jain, V. and Learned-Miller, E. (2010). Fddb: A benchmark for face detection in unconstrained settings. Technical Report UM-CS-2010-009, University of Massachusetts, Amherst.

Jones, M. J. and Rehg, J. M. (2002). Statistical color models with application to skin detection. *Int. J. Comput. Vision*, 46(1):81–96.

Khan, F. S., Anwer, R. M., van de Weijer, J., Bagdanov, A. D., Vanrell, M., and Lopez, A. M. (2012a). Color attributes for object detection. In *CVPR*, pages 3306–3313. IEEE.

Khan, F. S., van de Weijer, J., and Vanrell, M. (2012b). Modulating shape features by color attention for object recognition. *International Journal of Computer Vision*, 98(1):49–64.

Khan, R., van de Weijer, J., Khan, F. S., Muselet, D., Ducottet, C., and Barat, C. (2013). Discriminative color descriptors. In *CVPR*, pages 2866–2873. IEEE.

Li, H., Hua, G., Lin, Z., Brandt, J., and Yang, J. (2013). Probabilistic elastic part model for unsupervised face detector adaptation. In *The IEEE International Conference on Computer Vision (ICCV)*.

Mikolajczyk, K., Schmid, C., and Zisserman, A. (2004). Human detection based on a probabilistic assembly of robust part detectors. In *ECCV (1)*, volume 3021 of *Lecture Notes in Computer Science*, pages 69–82. Springer.

Ojala, T., Pietikäinen, M., and Mäenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(7):971–987.

Osuna, E., Freund, R., and Girosi, F. (1997). Training support vector machines: an application to face detection. pages 130–136.

Overett, G., Petersson, L., Brewer, N., Pettersson, N., and Andersson, L. (2008). A new pedestrian dataset for supervised learning. In *IEEE Intelligent Vehivles Symposium*, Eindhoven, The Netherlands.

Platt, J. C. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In *ADVANCES IN LARGE MARGIN CLASSIFIERS*, pages 61–74. MIT Press.

Romdhani, S., Torr, P., and Schölkopf, B. (2004). Efficient face detection by a cascaded support-vector machine expansion. *Royal Society of London Proceedings Series A*, 460:3283–3297.

Shechtman, E. and Irani, M. (2007). Matching local self-similarities across images and videos. In *IEEE Conference on Computer Vision and Pattern Recognition 2007 (CVPR'07)*.

Snoek, C. G. M. (2005). Early versus late fusion in semantic video analysis. In *In ACM Multimedia*, pages 399–402.

Terrillon, J.-C., Fukamachi, H., Akamatsu, S., and Shirazi, M. N. (2000). Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images. In *FG*, pages 54–63.

Van de Sande, K. E. A., Gevers, T., and Snoek, C. G. M. (2010). Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1582–1596.

Viola, P. and Jones, M. (2004). Robust real-time face detection. *International Journal of Computer Vision*, 57:137–154.

Wei, Y., Sun, J., Tang, X., and Shum, H.-Y. (2007). Interactive offline tracking for color objects. In *ICCV*, pages 1–8.

Weijer, J. V. D. and Schmid, C. (2006). Coloring local feature extraction. In *In ECCV, 2006. MENSINK et al.: TMRF FOR IMAGE AUTOANNOTATION*.

Wolf, L., Hassner, T., and Taigman, Y. (2008). Descriptor based methods in the wild. In *Real-Life Images workshop at the European Conference on Computer Vision (ECCV)*.

Zarit, B. D., Super, B. J., and Quek, F. K. H. (1999). Comparison of five color models in skin pixel classification. In *In ICCV99 Intl. Workshop on*, pages 58–63.

Zhang, L., Chu, R., Xiang, S., Liao, S., and Li, S. Z. (2007). Face detection based on multi-block lbp representation. In *Proceedings of the 2007 International Conference on Advances in Biometrics*, ICB'07, pages 11–18, Berlin, Heidelberg. Springer-Verlag.

Zhu, X. and Ramanan, D. (2012). Face detection, pose estimation, and landmark localization in the wild. In *CVPR*, pages 2879–2886.