Colour Invariants for Machine Face Recognition

Ognjen Arandjelović
Trinity College
University of Cambridge
Cambridge, CB2 1TQ
oa214@eng.cam.ac.uk

Roberto Cipolla
Department of Engineering
University of Cambridge
Cambridge, CB2 1PZ
cipolla@eng.cam.ac.uk

Abstract

Illumination invariance remains the most researched, yet the most challenging aspect of automatic face recognition. In this paper we investigate the discriminative power of colour-based invariants in the presence of large illumination changes between training and test data, when appearance changes due to cast shadows and non-Lambertian effects are significant. Specifically, there are three main contributions: (i) we employ a more sophisticated photometric model of the camera and show how its parameters can be estimated, (ii) we derive several novel colour-based face invariants, and (iii) on a large database of video sequences we examine and evaluate the largest number of colour-based representations in the literature. Our results suggest that colour invariants do have a substantial discriminative power which may increase the robustness and accuracy of recognition from low resolution images.

1. Introduction

In this paper we are interested in *colour-based* face representations for machine face recognition. Owing to its invariance to illumination [11] skin colour has been used extensively in detection and tracking applications e.g. hand tracking [10], face detection [6] and face segmentation [2]. In contrast, colour has received little attention in the face recognition community in spite of neurophysiological evidence that it is an important cue in recognition from low-resolution images [15].

The most complete comparison of colour-based representations for face recognition was published by Torres *et al*. [12] in which the discriminative power of different colour spaces (RGB, YUV and HSV) was evaluated. There are several limitations of the reported evaluation, which we address in this paper. Firstly, data with little illumination variation was used, as witnessed by the high recognition rate (85%) attained even using unprocessed luminance only.
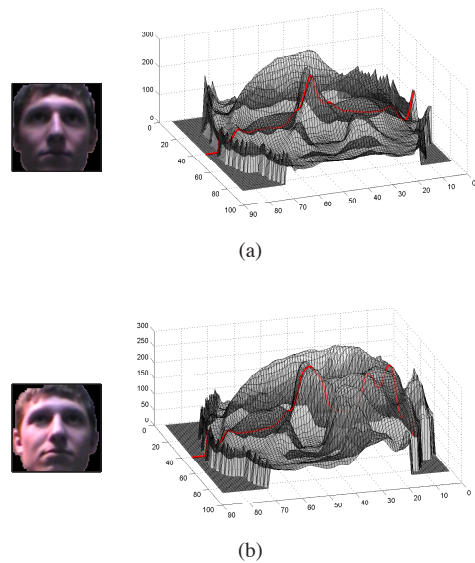

Figure 1. *The extent of facial appearance changes with illumination is difficult to fully appreciate as the human visual system is highly adapted to such variations. Above we visualize appearance as a 3D surface with height proportional to pixel intensity to illustrate the challenge posed to automatic recognition methods.*

In contrast, our data contains extreme illumination changes with prominent non-Lambertian effects. As we will demonstrate, these can have a dramatic effect on the recognition performance using different colour representations. Second, the implicitly employed photometric camera model is the simple linear model (see Sec. 2.1) which is in Sec. 3 shown to be less accurate than the more complex model proposed here. In this paper we also show how the parameters of the complex model can be recovered from face motion sequences, and describe and evaluate several illumination invariants based on it.
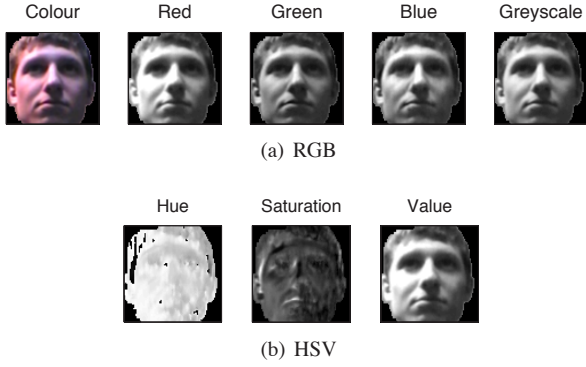
1

Figure 2. *(a) None of the channels in the RGB decomposition of an image are illumination invariant; (b) the Hue band of the HSV space shows the highest degree of invariance.*

2. Colour-Based Invariants

In this section we describe a detailed photometric model of a camera. Then, by first considering its simpler, special cases and working towards the most general form, we derive a number of colour-based illumination invariants. These are evaluated in Sec. 3.

2.1. Photometric camera model

Following several successful methods from the literature [9, 1], we too start from a weak photometric assumption that the measured intensity of a pixel is a linear function of the albedo $a(x,y)$ of the corresponding surface point:

$$I(x,y) = a(x,y) \cdot f(\Theta) \qquad (1)$$

where $\mathbf{f}$ is a function of illumination, shape and other parameters not modelled explicitly ($\Theta \equiv \Theta(x,y)$). We extend this approach for colour images by treating each of the colour channels $\mathbf{I}_C$ separately and describing surface albedo as dependent on the wavelength of incident light:

$$I_C(x,y) = a_C(x,y) \cdot f(\Theta) \qquad (2)$$

where channel C is either red (R), green (G) or blue (B), see Fig. 2.

In this paper we further augment this model to account for nonlinearities in the camera response. In particular, we include the (i) camera gamma parameter γ, (ii) linear gain G and (iii) the clipping, saturation function:

$$I_C(x,y) = \max\left[(a_C(x,y) \cdot f(\Theta))^{\gamma_C} \cdot G, \ 1.0 \right] \qquad (3)$$

see Fig. 3. Note that gamma is also light wavelength dependant, introducing three further unknowns: γ_R, γ_G and γ_B.
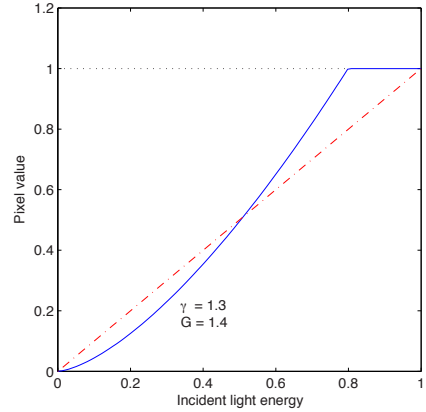

Figure 3. *The previously used linear photometric model of camera response significantly deviates from the more realistic, but also more complex model employed in this paper.*
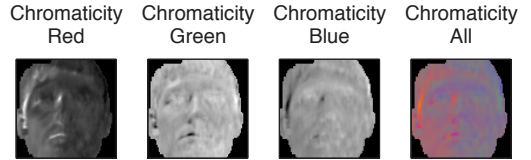

Figure 4. *Under the assumption of a unity gamma, chromaticity images are illumination invariant. While far less sensitive to illumination than the corresponding RGB channels in Fig. 2, there is still notable room for improvement. It is important to notice that the three chromaticity images exhibit different degrees of invariance, motivating our wavelength dependent gamma in (3).*

2.2. Unity gamma model and chromaticity

The simplest special case of the photometric camera model (3) that is of interest in this paper is obtained when $\gamma_R = \gamma_G = \gamma_B = 1.0$. In this case, the *chromaticity* images $\mathbf{H}_C$ ($C \in \{R, G, B\}$) are invariant both to illumination changes (i.e. to Θ in (3)) and camera parameters:

$$H_C(x,y) = \frac{I_C(x,y)}{\sum_{i \in \{R,G,B\}} I_i(x,y)} \qquad (4)$$

$$= \frac{a_C(x,y)}{\sum_{i \in \{R,G,B\}} a_i(x,y)} \qquad (5)$$

Examples are shown in Fig. 4.

2.3. Variable gamma model

An examination of chromaticity images shows both that they are generally not entirely invariant to illumination and also that red, green and blue chromaticities show *different degrees* of sensitivity. Referring back to our image formation model in (3), we can see that this means that the value

of gamma is not unity and furthermore, that it is wavelength dependent i.e. $\gamma_R \neq \gamma_G \neq \gamma_B$.

2.3.1 Estimating gammas

The key observation that we use to estimate the colour channels' gamma values (up to their ratio) is that faces have vertical symmetry. Consider an image $\mathbf{I}$ of a *frontal* face. We know that pixels (x_1, y) and (x_2, y) (where $x_2 = w+1-x_1$) correspond to surface points with the same shape and reflectance properties, see Fig. 5 (a). Then for non-saturated pixels it holds:

$$\forall C.\ I_C(x_1, y) = (a_C(x_1, y) \cdot f(\mathbf{\Theta}_1))^{\gamma_C} \cdot G \quad (6)$$

$$I_C(x_2, y) = (a_C(x_1, y) \cdot f(\mathbf{\Theta}_2))^{\gamma_C} \cdot G, \quad (7)$$

and eliminating $a_C(x_1, y)$:

$$\log \frac{I_C(x_1, y)}{I_C(x_2, y)} = \gamma_C \log \frac{f(\mathbf{\Theta}_1)}{f(\mathbf{\Theta}_2)}. \quad (8)$$

By applying (8) to two colour channels we can estimate the *ratio* of the two corresponding gammas, e.g. γ_R and γ_B:

$$\hat{\gamma}_R \equiv \frac{\gamma_R}{\gamma_B} = \log \frac{I_R(x_1, y)}{I_R(x_2, y)} \bigg/ \log \frac{I_B(x_1, y)}{I_B(x_2, y)}. \quad (9)$$

To increase the accuracy of the estimate in the presence of image noise and spatial discretization, we find the gamma value that achieves the best agreement across the entire image:

$$\hat{\gamma}_R \equiv \frac{\gamma_R}{\gamma_B} = \arg\min_{\gamma} \sum_{x_1} \sum_{y} \quad (10)$$

$$\left| \gamma \cdot \log \frac{I_B(x_1, y)}{I_B(x_2, y)} - \log \frac{I_R(x_1, y)}{I_R(x_2, y)} \right|^2. \quad (11)$$

Note that recovering gammas up to their ratio is the best that one can do without making further assumptions (such as imposing a prior on the wavelength dependent albedos) since a face with channel albedo $a_C(x, y)$ imaged by a camera with the corresponding gamma γ_C is indistinguishable from a face with the albedo $a_C(x, y)^{\gamma_C}$ imaged by a camera with a unity gamma.
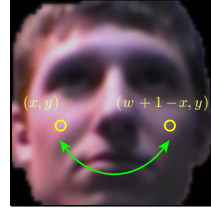
2.3.2 Unity gain camera

Let us first consider the special case of unity gain, i.e. $G = 1$ in (3). Defining semi-gamma normalized channels $\hat{\mathbf{I}}_C$ as

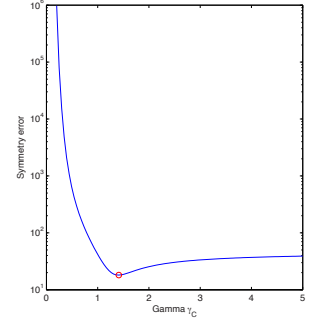$$\hat{I}_C(x, y) = [I_C(x, y)]^{\gamma_B / \gamma_C}, \quad (12)$$

it can be shown that:

$$\Phi_B(x, y) \equiv \frac{\log \hat{I}_R(x, y) - \log \hat{I}_B(x, y)}{\log \hat{I}_G(x, y) - \log \hat{I}_B(x, y)} \quad (13)$$

$$= \frac{\log a_R(x, y) - \log a_B(x, y)}{\log a_G(x, y) - \log a_B(x, y)}. \quad (14)$$

Figure 5. *(a) We use face symmetry to estimate the wavelength-dependent camera gamma. (b) The estimate is made by polling "votes" from all pixels and finding the gamma value that minimizes total disagreement across the image.*

The quantity $\mathbf{\Phi}_B$ is thus entirely independent of illumination or camera parameters (under the assumption of unity gain). Rather, it is a function of person-specific albedos $\mathbf{a}_C$ and is a colour-based invariant under the given photometric model. Similarly, so are $\Phi_R(x, y)$ and $\Phi_G(x, y)$:

$$\Phi_R(x, y) \equiv \frac{\log \hat{I}_G(x, y) - \log \hat{I}_R(x, y)}{\log \hat{I}_B(x, y) - \log \hat{I}_R(x, y)} \quad (15)$$

$$\Phi_G(x, y) \equiv \frac{\log \hat{I}_R(x, y) - \log \hat{I}_G(x, y)}{\log \hat{I}_B(x, y) - \log \hat{I}_G(x, y)} \quad (16)$$

We shall refer to $\Phi_R(x, y)$, $\Phi_G(x, y)$ and $\Phi_B(x, y)$ as, respectively, the red, green and blue-centred *log-Δ-ratios*.

2.3.3 Variable gain model

In Sec. 2.2 we considered a special case of the camera photometric model with unity gamma. We were able to derive two independent colour invariants by looking at each pixel individually. In the previous section we allowed gammas to vary. The increased number of unknown parameters meant that we could no longer find an invariant at each pixel – indeed, we used face symmetry as a further constraint. Now we consider the most general case of the model (3) in which *both* the camera gain and the wavelength dependent gammas are variable.

Much like before, we face the problem of having a higher number of unknowns than independent equations. In terms of our model, the ambiguity is posed by not being able to differentiate between a face with albedo $a_C(x, y)$ imaged by a camera with gain G and gamma γ_C, and a face with albedo $a_C(x, y) \cdot G^{1/\gamma_C}$ imaged by a camera with unity gain and gamma γ_C.

Consider the (say) blue-centred log-Δ-ratio introduced in (14), under the variable gamma/gain model:

$$\Phi_B(x,y;G) = \frac{\log \hat{I}_R(x,y) - \log \hat{I}_B(x,y)}{\log \hat{I}_G(x,y) - \log \hat{I}_B(x,y)} \qquad (17)$$

$$= \frac{\gamma_B \log \frac{a_R(x,y)}{a_B(x,y)} + \left(\frac{\gamma_B}{\gamma_R} - 1\right)\log G}{\gamma_B \log \frac{a_G(x,y)}{a_B(x,y)} + \left(\frac{\gamma_B}{\gamma_G} - 1\right)\log G}. \qquad (18)$$

Clearly Φ_B is now a function of the camera gain and thus no longer an invariant. However, if G was somehow known, the same invariant of (14) could be computed easily:

$$\frac{\log a_R(x,y) - \log a_B(x,y)}{\log a_G(x,y) - \log a_B(x,y)} = \qquad (19)$$

$$\frac{\log \hat{I}_R(x,y) - \log \hat{I}_B(x,y) - \left(\frac{\gamma_B}{\gamma_R} - 1\right)\log G}{\log \hat{I}_G(x,y) - \log \hat{I}_B(x,y) - \left(\frac{\gamma_B}{\gamma_G} - 1\right)\log G}. \qquad (20)$$

We use this by computing, and adjusting for, the *relative* camera gain between data sets when they are compared, and call this the *adaptive log-Δ-ratio*.

Consider two frontal faces from different sequences. If $\mathbf{\Phi_B}'$ is the blue-centred log-Δ-ratio of the reference, the relative camera gain is determined by minimizing:

$$\hat{G} = \arg\min_G \sum_x \sum_y$$

$$\left| \frac{\log \hat{I}_R(x,y)/\hat{I}_B(x,y) - \left(\frac{\gamma_B}{\gamma_R} - 1\right)\log G}{\log \hat{I}_G(x,y)/\hat{I}_B(x,y) - \left(\frac{\gamma_B}{\gamma_G} - 1\right)\log G} - \Phi_B'(x,y) \right|^2,$$
$$(21)$$

where non-primed variables correspond to the non-reference image.

It can be seen that the estimate of the relative gain, is accurate when the identity of the person in the compared data sets is the same. The value of $\hat{G}$ is not meaningful when the corresponding identities are different. This is however not a concern, as by the very nature of the invariant, in this case no camera gain will produce a good match.

2.4. Saturation, specular reflections and shadows

The final aspect of our camera model that we need to address is that of colour-wise "uninformative" pixels. We classify these into three groups: saturated, specular and shadowed.

Saturation is perhaps the easiest to understand as being uninformative: loss of information occurs as the energy of incident light is outside of the photo-sensor sensitivity range. In our photometric camera model the effects of saturation are represented by the clipping max function.

In contrast, within the context of this paper, intensely specular image regions are problematic not due to the limitations of practical imaging equipment, but rather due to

Figure 6. *Pixels detected as saturated (shown in red) are ignored.*

inherent physical reasons. This is because unlike isotropic, diffuse reflection, specular reflection by definition does not depend on surface albedo [7] and is effectively determined by the colour of incident light [13].

Finally, deeply shadowed pixels lack colour information because insufficient light was reflected to lend itself to wavelength/colour analysis. In the case of chromaticity, for example, this problem demonstrates itself through division by zero in (4).

Our approach. For simplicity, uninformative regions in this paper are excluded from consideration when appearance models are built (see Sec. 3.2). As a consequence, they do not contribute to the similarity score between sequences. We formally classify a pixel as uninformative if its luminance is either less than 3%, or more than 97% of the maximal luminance that can be represented, see Fig. 6.

Discussion. Before we proceed to the next section, we wish to add a brief clarification regarding "uninformative" pixels. Our claim is not that these are *entirely* lacking in discriminative information. As a simple example, if only a single colour channel is saturated, the remaining two channels can still be used to derive a colour constraint. Calling such image areas "less informative" would have probably been more appropriate, but we decided against it for the sake of avoiding awkward language constructs.

Also, we emphasize that we do not mean to suggest that these pixels are uninformative in *general*, but merely in the context of *colour-based invariants*. Indeed, spatial distribution of shadowed and specular pixels contains strong shape cues [4], amongst others.

3. Empirical Evaluation

The central premise of this paper is that in the treatment of colour for the purpose of face recognition, nonlinear effects in the photometric camera response are significant and need to be carefully modelled. In this section we first

Figure 7. *Cambridge Face Database contains extreme illumination conditions which also greatly vary between sequences. They are illustrated on a single frontal face for the purpose of isolating illumination effects only.*
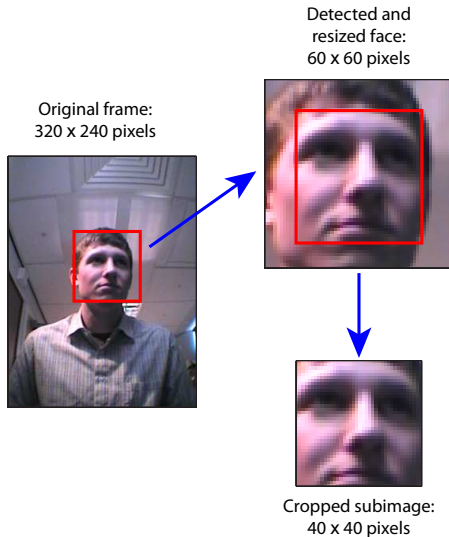

Figure 8. *Following detection, we automatically crop faces so as to eliminate any image regions which may interfere with the study of colour.*

present empirical evidence for this assertion and then quantify the contribution of each model parameter by evaluating the appropriate proposed colour-invariant.

3.1. Data

We conducted evaluation on a large database of face motion video sequences kindly provided to us by the *University of Cambridge* and described in detail in [1]. The 700 sequences in this database, each containing 100 frames, were acquired in a virtually unconstrained setting, thus making the recognition task representatively challenging for most practical applications. Specifically, the extent of illumination variation across the 7 different settings used for each of the 100 people, is illustrated in Fig. 7.

Faces, which vary in scale from (roughly) 40 to 80 pixels, were extracted from 320×240 pixel frames using the Viola-Jones detector [14]. They were then rescaled to the uniform scale of 60×60 pixels and cropped to the innermost 40×40 pixel subimage, as shown in Fig. 8.

3.2. Implementation details

Our aim in this evaluation was not necessarily to engineer the best performing system, but rather to obtain an assessment of *relative* performance of different representations and invariants. We chose *canonical correlations (CC)* between linear subspaces [5, 8] as a simple and well understood method for matching sets of fixed dimensionality vectors.

3.2.1 Set matching

Our basic algorithm for pairwise matching of face sets consists of two stages. Model estimation consists of fitting a linear subspace to each image set corresponding to a single input sequence. Two such sets are then compared and the first canonical correlation between the corresponding subspaces is used as the similarity measure. We now explain these steps in more detail.

Model estimation. Let $\mathbf{d}_i$ be a raster-ordered representation of the i-th detected face in a video sequence. The basis vectors of the corresponding linear subspace can be computed as the eigenvectors corresponding to the largest eigenvalue of the cross-correlation matrix $\mathbf{C} = \mathbf{D}\mathbf{D}^T/N$, where:

$$\mathbf{D} = \begin{bmatrix} \mathbf{d}_1 | \mathbf{d}_2 | \ldots | \mathbf{d}_N \end{bmatrix}. \qquad (22)$$

Model estimation with void elements. In Sec. 2.4 we explained why some image regions cannot be used to extract colour invariants. This means that the corresponding elements of $\mathbf{d}_i$ are undefined and PCA cannot be readily performed. We thus modify the basic model estimation algorithm to take this feature of our data into account.

Let $\mathbf{m}_i$ be the mask corresponding to $\mathbf{d}_i$, such that $m_i(j) = 0$ iff $d_i(j)$ is undefined. We then perform PCA on a modified cross-correlation matrix

$$\hat{\mathbf{C}} = \left(\mathbf{D}\mathbf{D}^T\right) \div \left(\mathbf{M}\mathbf{M}^T\right), \qquad (23)$$

where $\div$ denotes element-wise division and

$$\mathbf{M} = \begin{bmatrix} \mathbf{m}_1 | \mathbf{m}_2 | \ldots | \mathbf{m}_N \end{bmatrix}. \qquad (24)$$

CC matching. The first canonical correlation between two subspaces spanned by bases $\mathbf{B}_1$ and $\mathbf{B}_2$ can be computed as the largest singular value of the matrix $\mathbf{B}_1^T\mathbf{B}_1$ [3]. It it equal to the cosine of the smallest angle between vectors of the two spaces.

3.2.2 Detection and synthesis of frontal faces

The extraction of colour invariants proposed in Sec. 2.3 relies on the availability of an image of a frontal face to recover a set of camera parameters. At the very least, this means that we need a reliable way of automatically selecting the frontal-most face from the pool of all detections in a sequence, or more likely, an algorithm for *synthesizing* a frontal face from non-frontal detections. We summarize our approach:

1. map all face images $\mathbf{I}$ onto a quasi-illumination invariant domain $\mathbf{I} \rightarrow \mathbf{S}$

2. quantify the degree of "frontality" of maps of all detected faces

3. select the frontal-most (highest "frontality") face $\mathbf{I}_f$:

    (a) from the neighbourhood of the $\mathbf{S}_f$ estimate the 2D plane tangential to the appearance manifold

    (b) perform extrapolation from $\mathbf{S}_f$ along the tangent plane to the point $\mathbf{S}'_f$ nearest to the *vertical symmetry hyperplane*

    (c) inverse map $\mathbf{S}'_f$ to $\mathbf{I}'_f$

4. result: $\mathbf{I}'_f$ is a synthetic image of the frontal face

Measuring "frontality". To find the face in a data set which is closest to frontal, we need a way of quantifying the degree of face "frontality". Our approach consists of computing a distance transformed edge map of each face image, which is a quasi-illumination invariant representation, and then measuring the cosine of the angle between its rasterized left and (mirrored) right halves. Fig. 9 (a) illustrates the basic principle, while Fig. 9 (b) shows typical responses to differently oriented faces.

Finding the inverse map. After localizing the frontal-most face, we use the distribution of its neighbours to extrapolate in the direction tangential to the appearance manifold, see Fig. 9 (c). Since the face detector normalizes for face scale and location, the two dominant modes of appearance changes in a singe data set correspond to varying pitch and yaw, thus resulting in a 2D manifold. We perform extrapolation in the quasi-illumination invariant space of distance transformed edge maps, maximizing vertical symmetry. The result in the appearance domain is then obtained by linearly combining the corresponding appearance images.

3.3. Results and discussion

We summarized our experimental results in Tab. 1. Firstly, note the grouping of different representations in the table into two categories: colour space transformations and

Table 1. *A summary of experimental results.*

Representation	Recognition rate (%)	
Greyscale	64.6	Colour space transformations
Colour channels, all (RGB)	64.9	
Colour channel, red	55.5	
Colour channel, green	66.5	
Colour channel, blue	67.9	
HSV	66.1	
Saturation	56.2	
Value	65.8	
Hue	35.1	Colour-based invariant signatures
Chromaticity, red	48.5	
Chromaticity, green	56.8	
Chromaticity, blue	51.3	
Chromaticity, all	58.2	
Log-Δ-ratio, red-centred	39.2	
Log-Δ-ratio, green-centred	57.8	
Log-Δ-ratio, blue-centred	57.8	
Log-Δ-ratio, all	64.5	
Adaptive log-Δ-ratio, red-centred	40.3	
Adaptive log-Δ-ratio, green-centred	64.9	
Adaptive log-Δ-ratio, blue-centred	63.7	
Adaptive log-Δ-ratio, all	65.1	

colour-based invariant signatures. The representations of the former group, while functions of colour, are also inherently dependent on the manner in which a face is illuminated. On the other hand, the representations of the latter group are all invariants, each under a specific photometric model.

The results obtained using raw images are useful as a benchmark for quantifying the severity of illumination variation in the database. Specifically, in comparison to Torres *et al.* [12], our data set is far more challenging with approximately 25% lower recognition rate obtained using unprocessed greyscale. This difference is even more significant when it is taken into account that we performed recognition from video sequences, thus using more data and effectively normalizing for pose, as well as that our matching algorithm is more sophisticated in comparison to the simple PCA in
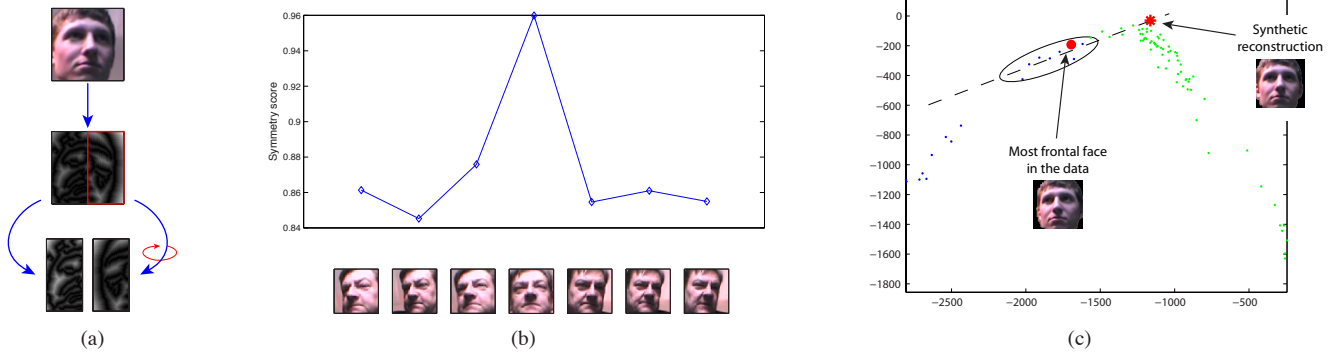
Figure 9. *(a,b) We quantify the degree of face "frontality" by measuring vertical symmetry of distance transformed Canny edges. (c) After finding the frontal-most face in a data set we use its neighbourhood to synthetically improve the result by performing 2D linear extrapolation. For clarity, this is illustrated in the 2D principal component space (blue points represent face images in a hypothetical data set; green points represent samples from the remainder of the actual appearance manifold, which can be used to verify the accuracy of synthetic frontal face reconstruction).*

[12].

Much like Torres *et al.*, we too found no statistically significant improvement when using the RGB colour space, either a single channel at a time, or all together. However, as the remainder of our results will show, we argue that it would not be correct to conclude from this (as Torres *et al.* do) that colour information has nothing to add to the discriminative power of luminance.

It is interesting to note that in contrast to Torres *et al.* who found the three colour components equally informative, in our experiments Red was notably worse than Green and Blue. The same was found in the case of chrominance components. For this reason we examined the three channels in more detail, see Fig. 10. Red was found to be the greatest in magnitude, which is not surprising given the red-dominant colour of skin. Interestingly, the correlation between Green and Blue was consistently quite high, and quite low (but very variable as suggested by the variance) between Red and either Green or Blue.

Superficially, the recognition results achieved using individual HSV components may seem somewhat surprising: the performance of the (near) invariant Hue (see Sec. 2.2) is rather disappointing, with the heavily illumination-affected Value correctly matching twice as many individuals. The performance of Hue is indicative of the inherent discriminative power of *pure* colour. In effect, it is this performance that we are set on scrutinizing and improving upon in this paper. It is also insightful to consider why Value performed so relatively well, in the light of the widely accepted claim that illumination is one of the foremost challenges to face recognition. Briefly put, the reason is that it is the large *changes* in illumination that present difficulties; shadows and highlights can in fact help discern between individuals,

effectively by placing constraints on the head shape, otherwise lost in the process of projection onto the image plane.

Recognition rate attained using individual chromaticity components significantly exceeded that of Hue and in combination nearly matching greyscale performance ("Chromaticity, all"). This supports our main premise and the proposed photometric model: by analyzing the dependence of measured RGB values at each pixel on camera gain, we were able to derive a representation that is not affect by gain changes, see Sec. 2.2.

Our introduction of wavelength-dependent gammas in Sec. 2.3 provides further substantial improvement, adaptive log-Δ-ratios expectedly performing better than simple log-Δ-ratios (see Sec. 2.3.2 and 2.3.3). The attained rate slightly exceeds that of the greyscale representation (as well as RGB), which is quite remarkable given that the log-Δ-ratios are pure colour invariants and thus *complementary* to greyscale. These results suggest that colour is in fact much more promising for face recognition than previously acknowledged.

Interestingly, despite the very different nature of the log-Δ-ratios based representations and chromaticity or RGB components, the representation corresponding to the Red channel was found to be consistently worse in all cases than those corresponding to Green or Blue. We found no satisfying explanation for this and suggest that more research is needed.

4. Conclusion

This paper analyzed the importance of colour in machine recognition of faces. It was argued and experimentally demonstrated that the previously largely ignored nonlinear effects in the photometric response of the camera are
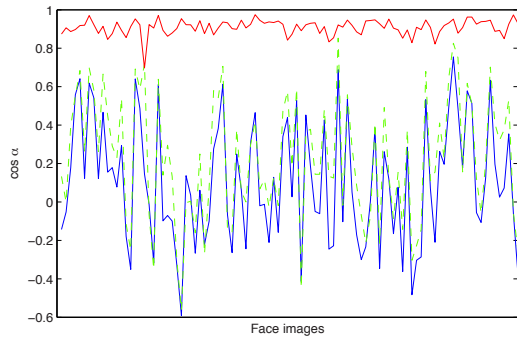
(a)

Angle	Mean	Variance
Red-Green	0.1063	0.3000^2
Red-Blue	0.2272	0.3026^2
Green-Blue	0.9077	0.0398^2

(b)

Magnitude (/ luminance)	Mean	Variance
Red	1.1141	0.0404^2
Green	0.8766	0.0209^2
Blue	1.0149	0.0233^2

(c)

Figure 10. *Pairwise angles between RGB colour components (from the mean luminance) (a) plotted across a portion of our face data set and (b) the mean statistics for the entire database, and (c) their magnitudes (relative to luminance).*

in fact substantial and should be modelled. Thus, a number of novel colour invariants were developed for several models with different complexities. Their recognition performance on a large database with extreme illumination variability suggests that the use of colour may significantly improve greyscale-based matching algorithms.

We believe that the reported results open a number of promising areas for further work. The most immediate research direction we intend to pursue, motivated by the success of similar methods in matching greyscale appearance, is that of developing algorithms which better exploit the manifold structure of colour-based invariant representations.

References

[1] O. Arandjelović and R. Cipolla. Face recognition from video using the generic shape-illumination manifold. *In Proc. European Conference on Computer Vision (ECCV)*, 4:27–40, 2006. 2, 5

[2] O. Arandjelović, G. Shakhnarovich, J. Fisher, R. Cipolla, and T. Darrell. Face recognition with image sets using manifold density divergence. *In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1:581–588, 2005. 1

[3] Å. Björck and G. H. Golub. Numerical methods for computing angles between linear subspaces. *Mathematics of Computation*, 27(123):579–594, 1973. 5

[4] A. Blake and G. Brelstaff. Geometry from specularities. *In Proc. IEEE International Conference on Computer Vision (ICCV)*, pages 394–403, 1988. 4

[5] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28:321–372, 1936. 5

[6] R.-L. Hsu, M. Abdel-Mottaleb, and A. K. Jain. Face detection in color images. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 24(5):696–706, 2002. 1

[7] S. K. Nayar, K. Ikeuchi, and T. Kanade. Identification of human faces based on isodensity maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 13(7):611–634, 1991. 4

[8] E. Oja. *Subspace Methods of Pattern Recognition*. Research Studies Press and J. Wiley, 1983. 5

[9] T. Riklin-Raviv and A. Shashua. The quotient image: Class based re-rendering and recognition with varying illuminations. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 23(2):219–139, 2001. 2

[10] B. Stenger, A. Thayananthan, P. H. S. Torr, and R. Cipolla. Filtering using a tree-based estimator. *In Proc. IEEE International Conference on Computer Vision (ICCV)*, 2:1063–1070, 2003. 1

[11] M. Störring, H. J. Andersen, and E. Granum. Skin colour detection under changing lighting conditions. *Symposium on Intelligent Robotics Systems*, pages 187–195, 1999. 1

[12] L. Torres, J. Y. Reutter, and L. Lorente. The importance of color information in face recognition. *In Proc. IEEE International Conference on Image Processing (ICIP)*, 1999. 1, 6, 7

[13] S. Umeyama and G. Godin. Separation of diffuse and specular components of surface reflection by use of polarization and statistical analysis of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 26(5):639–647, 2004. 4

[14] P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision (IJCV)*, 57(2):137–154, 2004. 5

[15] A. Yip and P. Sinha. Role of color in face recognition. *Perception*, 31(5):995–1003, 2002. 1