Research Article

Benefits of Using Decorrelated Color Information for Face Segmentation/Tracking

Mircea C. Ionita and Peter Corcoran

Department of Electronic Engineering, National University of Ireland, Galway, Nun's Island, Galway, Ireland

Correspondence should be addressed to Mircea C. Ionita, mionita@alpha.imag.pub.ro

We analyze in this paper the benefits that can be derived from employing color image alignment techniques in the context of face segmentation or tracking based on texture (defined as the patch of intensities) template matching. By making full use of the decorrelated color information, improvements on the accuracy of the segmentation are demonstrated. This is intended to enhance the face segmentation algorithm by increasing its robustness to differences in images caused by various image acquisition devices or settings or by variations in the ambient illumination conditions.

1. INTRODUCTION

The use of color information becomes increasingly important in nowadays image processing applications as inexpensive color image acquisition devices become easily available. Color image processing permits a more extensive image representation, which expectedly leads to better results.

We deal in this paper with the specific case of face segmentation which employs face modeling techniques. This can also be viewed in the more general context of deformable template matching, using for this purpose a statistical model of shape variations. Extensive work has been carried out in the area of face modeling and face segmentation using statistical models [1–5]. These techniques have initially been developed for gray level images. Extensions have later been proposed for color images [6, 7]. Some advantages of using the color extension have been demonstrated, but mostly for working in a controlled image acquisition environment. Processing color information can thus be challenging, especially when designing more general applications that are supposed to work within unconstrained image acquisition conditions. We demonstrate in this paper some positive results when using the decorrelated color information for applications which include face segmentation and face tracking, intended to work under no predefined constrains.

The outline of this paper is as follows. In Section 2, we briefly describe several decorrelated color spaces in terms of their transforms from the common *RGB* color space; we also include a comparison between these color spaces in terms of how well they are able to decorrelate image channels on a series of test images. In Section 3, a face segmentation method is described, based on a statistical shape model and a fixed face texture template. The limitations of the application described in Section 3 are addressed in Section 4, introducing some texture alignment and color transfer techniques in order to adapt the texture template to the color distribution of the current image. These operations are facilitated by converting the texture data to one of the decorrelated color spaces presented in Section 2. In Section 5, we show our experiments performed on a general face image database; the database is built as a mixture of images, gathered mostly from various standard face image and video databases; a set of comparative results is provided in Section 5. Finally, in Section 6 we draw the conclusions of our work.

2. IMAGE DECORRELATION WITH RESPECT TO COLOR INFORMATION

Colorwise image decorrelation is useful for applying color image processing operations independently on each image channel.

2.1. Karhunen-Loève transform

The *Karhunen-Loève transform (KLT)* is optimal in terms of energy compaction and mean-squared error minimization for a truncated representation. By applying KLT to a color image, it creates image basis vectors which are orthogonal, and it thus achieves complete decorrelation of the image channels [8–12] as follows:

$$\mathbf{C}_{RGB} = \mathbf{A} \cdot (\mathbf{I} - \boldsymbol{\mu}_I), \tag{1}$$

where $\mathbf{I} = [I_R, I_G, I_B]^T$ contains the image color signals and $\boldsymbol{\mu}_I = [\mu_R, \mu_G, \mu_B]^T$, with $\mu_x = E[x]$; E is the mathematical expectation. $\mathbf{C}_{RGB}$ is the covariance matrix of the of the image color signals as follows:

$$\mathbf{C}_{RGB} = \begin{bmatrix} C_{RR} & C_{RG} & C_{RB} \\ C_{GR} & C_{GG} & C_{GB} \\ C_{BR} & C_{BG} & C_{BB} \end{bmatrix} \tag{2}$$

with $C_{xy} = E[(x - E[x])(y - E[y])]$, $x, y \in \{R, G, B\}$. $\mathbf{A}$ is the transformation matrix formed by the eigenvectors of the covariance matrix $\mathbf{C}_{RGB}$:

$$\mathbf{A} = \begin{bmatrix} \mathbf{k}_1 & \mathbf{k}_2 & \mathbf{k}_3 \end{bmatrix}^T = \begin{bmatrix} k_{11} & k_{21} & k_{31} \\ k_{12} & k_{22} & k_{32} \\ k_{13} & k_{23} & k_{33} \end{bmatrix}^T. \tag{3}$$

Yet, KLT is data dependant, meaning that it requires the recalculation of the transformation matrix $\mathbf{A}$ for each set of data (e.g., each new image).

2.2. $I_1 I_2 I_3$ color space

An interesting color space is $I_1 I_2 I_3$, proposed by Ohta et al. [13], which realizes a statistical minimization of the interchannel correlations (decorrelation of the *RGB* components) for natural images. The conversion from *RGB* to $I_1 I_2 I_3$ is given by the simple linear transformation in (4) as follows:

$$\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & -1/2 \\ -1/4 & 1/2 & -1/4 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \tag{4}$$

I_1 stands as the achromatic (intensity) component, while I_2 and I_3 are the chromatic components. We remark that the simple numeric transformation from *RGB* to $I_1 I_2 I_3$ enables simple and efficient transformation of datasets between these two color spaces.

$I_1 I_2 I_3$ was designed as an approximation for the KLT of the *RGB* data to be used for region segmentation on color images. As the transformation to $I_1 I_2 I_3$ represents a good approximation of the KLT for a large set of natural images, the resulting color channels are almost completely decorrelated.

In the previous work of Ohta et al., the discriminating power of 109 linear combinations of *R*, *G*, and *B* was tested on eight different color scenes. The selected linear combinations were gathered such that they could successfully be used for segmenting important (large area) regions of an image, based on a histogram threshold. It was found that 82 of the linear combinations had all positive weights, corresponding mainly to an intensity component which is best approximated by I_1; another 22 showed opposite signs for the weights of *R* and *B*, representing the difference between the *R* and *B* components which are best approximated by I_2; finally, the remaining 4 linear combinations could be approximated by I_3. Thus it was shown that the I_1, I_2, and I_3 components in (4) are effective for discriminating between different regions and that they are significant in this order [13]. We can further conclude, based on the above figures, that the percentage of color features which are well discriminated on the first, second, and third channels is around 76.15%, 20.18%, and 3.67%, respectively.

$I_1 I_2 I_3$ is also found in [14] to perform better as compared to other color space implementations like YIQ, CIELAB, and *UVW* for segmentation of color images based on *Markov random field (MRF)* processing. In [15], the $I_1 I_2 I_3$ color space was used for color image segmentation based on an MRF model and simulated annealing due to its effectiveness in terms of the quality of the segmentation and the reduced complexity of the transformation.

2.3. $l\alpha\beta$ color space

Assuming that the human visual system is ideal for processing natural scenes, Ruderman et al. [16] developed the $l\alpha\beta$ color space, which also minimizes the correlation between channels for natural images. The conversion from *RGB* is realized by means of an initial transform to *LMS* cone space, followed by a conversion of the data to logarithmic space (used to reduce skewness):

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.3811 & 0.5783 & 0.0402 \\ 0.1967 & 0.7244 & 0.0782 \\ 0.0241 & 0.1288 & 0.8444 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \tag{5}$$

$$\begin{matrix} \mathbf{L} & & \log L \\ \mathbf{M} & = & \log M \\ \mathbf{S} & & \log S \end{matrix}. \tag{6}$$

Finally, the $l\alpha\beta$ data is obtained from

$$\begin{bmatrix} l \\ \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} & 0 & 0 \\ 0 & 1/\sqrt{6} & 0 \\ 0 & 0 & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -2 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{L} \\ \mathbf{M} \\ \mathbf{S} \end{bmatrix}. \tag{7}$$

This color space has successfully been used in [17, 18] for image color transfer operations, which will be described in Section 4.1.

2.4. Comparison between the different color image representations

The correlation between two image channels is given by

$$\rho_{ij} = \frac{E[(v_i - E[v_i])(v_j - E[v_j])]}{\sqrt{E[(v_i - E[v_i])^2] E[(v_j - E[v_j])^2]}}, \tag{8}$$

TABLE 1: Decorrelation results over the tested database.

Image representation	Correlation coefficients (mean/std)			
	ρ_{12}	ρ_{13}	ρ_{23}	R_{total}
RGB	0.93/0.06	0.84/0.12	0.95/0.07	0.91/0.07
KLT	0/0	0/0	0/0	0/0
$I_1 I_2 I_3$	0.29/0.22	0.25/0.21	0.52/0.26	0.35/0.15
$l\alpha\beta$	0.32/0.20	0.27/0.20	0.74/0.19	0.44/0.13

FIGURE 1: General, mixed color face image database.

where v_i and v_j represent the ith and jth image channel signals, respectively, (with $i, j = \overline{1,3}$, $i \neq j$), for a certain color image representation.

The total interchannel correlation is calculated as follows:

$$R_{\text{total}} = \frac{|\rho_{12}| + |\rho_{13}| + |\rho_{23}|}{3}. \qquad (9)$$

The correlation coefficients have been measured for several test images (see Figure 1) in the discussed color image representations using the above formulae [19]. Results are summarized in Table 1.

It can be observed that the *RGB* representation presents a very high interchannel correlation, while the $I_1 I_2 I_3$ and $l\alpha\beta$ image representations significantly reduce this correlation. As stated above, the KLT, which is adapted to each particular image, achieves total decorrelation of the image channels.

3. FACE SEGMENTATION USING DEFORMABLE TEMPLATE MATCHING

Note that the term *texture*, frequently used in this paper, refers in the context of this work to the set of pixel intensities across an object, also subsequent to a suitable normalization.

3.1. Statistical shape models

We are interested in designing a shape model robust to head pose variations. The shape is defined as the set of positions of some fiducial points on the face. The model is statistically built from a training dataset which contains image examples, annotated with a fixed set of landmark points. The sets of 2-D coordinates of the landmark points define the shapes inside the image frame. These shapes are aligned using the generalized *procrustes analysis* [20], a technique for removing the differences in translation, rotation, and scale between the training set of shapes. This defines the shapes in the normalized frame.

Let N be the number of training examples. Each shape example is represented as a vector s of concatenated coordinates of its points $(x_1, x_2, \ldots, x_L, y_1, y_2, \ldots, y_L)^T$, where L is the number of landmark points. *principal components analysis (PCA)* is then applied to the set of aligned shape vectors reducing the initial dimensionality of the data. It can be noted that PCA is very similar to KLT. In a geometric interpretation, KLT can be viewed as a rotation of the coordinate system, while for PCA, the rotation of the coordinate system is preceded by a shift of the origin to the mean point [21]. Shape variability is thus linearly modeled as a base (mean) shape plus a linear combination of shape eigenvectors:

$$\mathbf{s}_m = \bar{\mathbf{s}} + \Phi_s \mathbf{b}_s, \qquad (10)$$

where $\mathbf{s}_m$ represents a modeled shape, $\bar{\mathbf{s}}$ is the mean of the aligned shapes, $\Phi_s = (\phi_{s_1}|\phi_{s_2}|\cdots|\phi_{s_p})$ is a matrix having p shape eigenvectors as its columns ($p < N$); finally, $\mathbf{b}_s$ defines the set of parameters of the shape model. p is chosen so that a certain percentage of the total variance of the data is retained.

The standard deviation for each parameter of the face model, as resulted from the training dataset, provides its dynamic range. By altering the model parameters within their dynamic range helps insuring that only plausible instances of the modeled object are being generated. A description of the way in which the optimal model parameters for a new image can automatically be estimated follows in Section 3.2.

3.2. Face texture template optimization algorithm

In order to optimize the face model parameters, a texture template is also required. The separation between shape and texture is realized using a reference shape. Based on this reference shape, the so-called *texture examples* can be extracted. The reference shape is usually chosen as the pointwise mean of the shape examples. The texture examples are defined in the normalized frame of the reference shape. Each image example is then distorted such that the points that define its attached shape, used as control points, match the reference shape, such that the topology is preserved. An image warping method is employed for this purpose. Image warping methods are discussed in Section 3.3.

Subsequent to the warping stage, all shape differences between the image examples have been removed. The texture across each image object is thus mapped into a shape-normalized representation. The resulting images are also called the image examples in the normalized frame. For each of these images, the corresponding pixel values across their common shape are scanned to form the texture vectors $\mathbf{t}_{im} = (t_{im_1}, t_{im_2}, \ldots, t_{im_P})^T$, where P is the number of texture samples.

Based on previous experiments, we remark that the variability of the shape component of the face is much more important than the variability of the texture component in terms of a successful segmentation of the face. Due to this fact, we consider in the following a simplified formulation of a model-based face segmentation technique, where the modeled image is represented by a fixed texture template; extensions could be made so that to include texture variability, yet that was beyond the purpose of the current work. Thus during an optimization stage (fitting the model to a query image), the parameters to be found are $\mathbf{p} = \begin{pmatrix} \mathbf{g}_s \\ \mathbf{b}_s \end{pmatrix}$, where $\mathbf{g}_s$ are the shape 2-D position, 2-D rotation, and scale parameters inside the image frame, and $\mathbf{b}_s$ are the shape model parameters. The optimization of the parameters $\mathbf{p}$ is realized by minimizing the reconstruction error between the query image and the modeled image. The error is evaluated in the coordinate frame of the model, that is, in the normalized texture reference frame, rather than in the coordinate frame of the image. The difference between the query image and the modeled image is thus given by the difference between the (normalized) image texture and the (normalized) template texture as follows:

$$\mathbf{r}(\mathbf{p}) = \mathbf{t}_s - \mathbf{t}_t, \tag{11}$$

and $\|\mathbf{r}(\mathbf{p})\|^2$ is the reconstruction error, with $\|\cdot\|$ marking the Euclidean norm.

A first order Taylor extension of $\mathbf{r}(\mathbf{p})$ is given by

$$\mathbf{r}(\mathbf{p} + \delta\mathbf{p}) \simeq \mathbf{r}(\mathbf{p}) + \frac{\partial \mathbf{r}}{\partial \mathbf{p}} \delta\mathbf{p}. \tag{12}$$

$\delta\mathbf{p}$ should be chosen so that to minimize $\|\mathbf{r}(\mathbf{p} + \delta\mathbf{p})\|^2$. It follows that:

$$\frac{\partial \mathbf{r}}{\partial \mathbf{p}} \delta\mathbf{p} = -\mathbf{r}(\mathbf{p}). \tag{13}$$

Normally, the gradient matrix $\partial\mathbf{r}/\partial\mathbf{p}$ should be recomputed at each iteration. Yet, as the error is estimated in a normalized texture frame, it was shown that this gradient matrix may be considered as fixed, being thus possible to precompute it from a training dataset; these techniques, introduced in [22], and extended so that to also incorporate a statistical texture variation model (as opposed to a fixed texture template described above), are called *active appearance models (AAMs)*. Using this technique, each parameter in $\mathbf{p}$ is systematically displaced from its known optimal value retaining the normalized texture differences. The resulted matrices are then averaged over several displacement amounts and over several training images. The update direction of the model parameters $\mathbf{p}$ is then given by

$$\delta\mathbf{p} = -\mathbf{R}\mathbf{r}(\mathbf{p}), \tag{14}$$

where $\mathbf{R} = ((\partial\mathbf{r}/\partial\mathbf{p})^T(\partial\mathbf{r}/\partial\mathbf{p}))^{-1}(\partial\mathbf{r}/\partial\mathbf{p})^T$ is the pseudoinverse of the determined gradient matrix, which can be precomputed as part of the training stage. The parameters $\mathbf{p}$ continue to be updated iteratively until the error can no longer be reduced and convergence is declared.

3.3. A TPS-based model fitting technique

Piecewise affine warping is extensively used in techniques like AAM due to its reduced computational costs. A triangulation (e.g., Delauney) is used to partition the convex hull of the control points. The points inside triangles are then mapped via an affine transformation which uniquely assigns the corners of a triangle to their new positions. Although the assumption that the face patches are piecewise affine within the triangles is a satisfactory solution when there is a sufficiently large number of landmark points, it also shows an important drawback. This refers to the fact that, when modeling large face pose variations, corners of some triangles tend to get reversed due to occlusions of the corresponding landmark points. This obviously affects the image warping outcome by creating erroneous face patches. The errors are further propagated into the fitting algorithm, resulting in an incorrect fit. That is why the piecewise warping method works well mostly for modeling frontal or nearly frontal faces.

A more advanced and accurate warping method is obtained by employing the *thin plate splines (TPSs)*, introduced in [23]. A short description of this warping method is also given in the appendix. An initial drawback of using the thin plate splines was represented by the fact that they were quite expensive to calculate. The solution requires the inversion of a $p \times p$ matrix (the bending energy matrix) which has a computational complexity of $O(N^3)$, where p is the number of points in the dataset (i.e., the number of pixels in the image); furthermore, the evaluation process is $O(N^2)$. Fortunately, important progress has been made in order to speed this process up. An approximation approach was proved in [24] to be very efficient in dealing with the first problem, reducing greatly the computational burden. As far as the evaluation process is concerned, the *multilevel fast multipole method (MLFMM)* framework was described
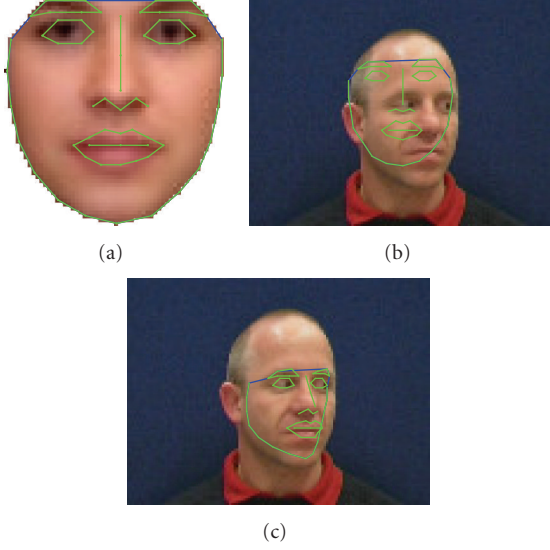
FIGURE 2: Face modeling. (a) Original texture template; (b) model initialization; and (c) converged model.
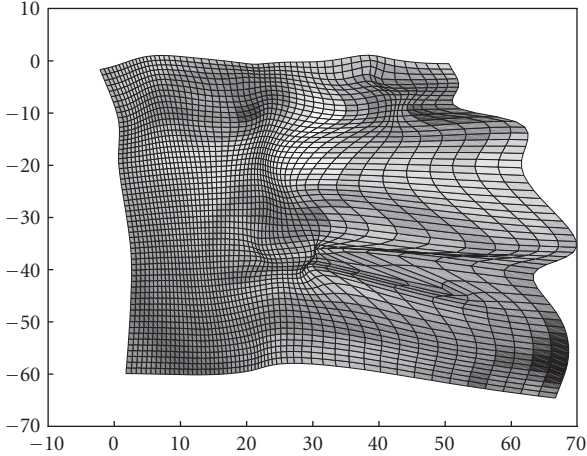

FIGURE 3: TPS-based convergence—TPS-deformed grid for the rectangular region around the fitted shape.

in [25] for the evaluation of two-dimensional polyharmonic splines, while in [26] this work was extended for the specific case of TPS, showing that a reduction of the computational complexity from $O(N^2)$ to $O(N \log N)$ is indeed possible. Thus the computational difficulties involving the use of TPS have been to an important extent removed.

We show in Figures 2 and 3 an example of fitting the model based on TPS warping. The error is evaluated relative to the number of available data points after the deformation.

4. IMPROVED MODEL FITTING BY MEANS OF LOCAL COLOR TRANSFER

A face detection algorithm is firstly applied for the current image. We used here the Viola-Jones face detector [27], which is based on the AdaBoost algorithm [28]. A statistical relation between the face detector estimates for the face position and size (rectangle region) and the position and size of the reference shape inside the image frame is initially learnt (offline) from a set of training images. This relation is then used to obtain a more accurate initialization for the reference shape, tuned with the employed face detection algorithm. It is also important to have a reasonably close initialization to the real values in order to insure the convergence of the fitting algorithm described in Section 3. Color statistics are then extracted across the convex hull of landmark points of the initialized reference shape.

4.1. Image color transfer

According to [17], color can be transferred between two images (global color transfer) using the formula in (15), applied in the $l\alpha\beta$ color space:

$$g(c_t) = \mu_s + \frac{\sigma_s}{\sigma_t}(c_t - \mu_t), \qquad (15)$$

where μ and σ are, respectively, the mean and standard deviation of the Gaussian distribution in the considered color space.

For local color transfer between two images, color statistics (e.g., mean and variance of the Gaussian-modeled color distribution) are gathered from the target and source image, respectively, and used to calculate the *color influence map (CIM)*. CIM contains the weights for each pixel in the target image, determined based on their proximity to the color range in the source image.

Consider the distance between a pixel and the center of the color distribution. For three-dimensional color data this is the Mahalanobis distance given by

$$d_M(\mathbf{c}, \boldsymbol{\mu}) = \sqrt{(\mathbf{c} - \boldsymbol{\mu})\mathbf{S}^{-1}(\mathbf{c} - \boldsymbol{\mu})}, \qquad (16)$$

where $\mathbf{S}$ is the covariance matrix of the three-variate color texture vector.

Yet, if a decorrelated color space is used, then the covariance matrix $\mathbf{S}$ is close to being diagonal and (16) reduces to the *normalized Euclidean distance* (17):

$$d(\mathbf{c}, \boldsymbol{\mu}) = \left\| \frac{\mathbf{c} - \boldsymbol{\mu}}{\boldsymbol{\sigma}} \right\|^2, \qquad (17)$$

where $\boldsymbol{\sigma}$ is the standard deviation vector of $\mathbf{c}$ over the sample set.

The weights in CIM are calculated using a function of the above distance $f(d)$, for which the following conditions should be met as follows:

$$\lim_{d \to \infty} f(d) = 0; \qquad f(0) = 1. \qquad (18)$$

The function below was proposed in [18] to be used with the $l\alpha\beta$ color space:

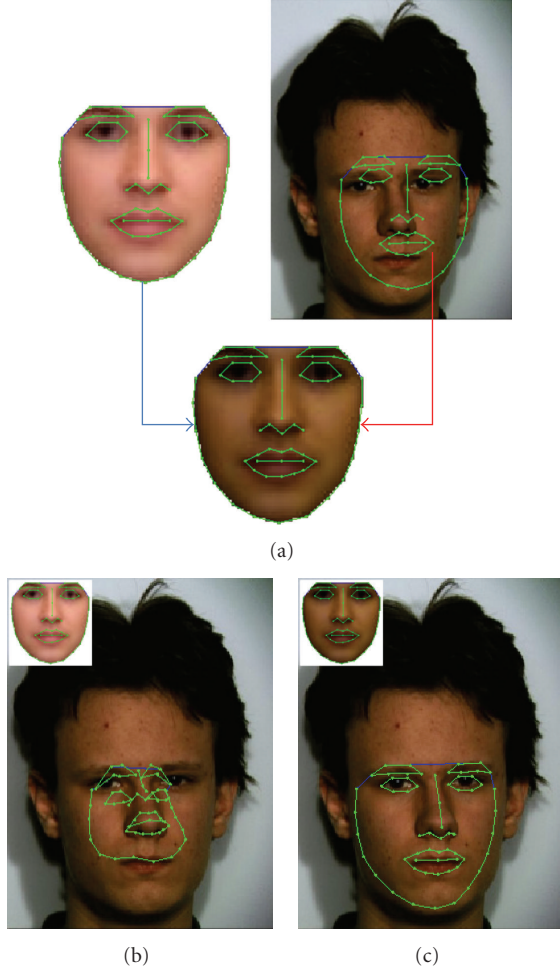$$f(d) = e^{-3d^2}. \qquad (19)$$

FIGURE 4: Face modeling using color transfer. (a) Generating the color-adapted template from the original template and the estimated face region in the image; (b) converged model with no color transfer; and (c) converged model using the color-adapted template.

The color transfer equation in (15) was also extended in [18] to

$$g(c_t) = c_t + f(d)\left(\mu_s + \frac{\sigma_s}{\sigma_t}(c_t - \mu_t) - c_t\right), \qquad (20)$$

or, if a single color **c** is used as source for color transfer,

$$g(c_t) = c_t + f(d)(\mathbf{c} - \mu_t). \qquad (21)$$

4.2. Adaptive texture template matching

Using a decorrelated color space (see Section 2), the color of the texture template (see Figure 4(a)) can be adapted to the current image, increasing the chance of a correct fitting (correct-face segmentation) of the face model. Experimental results to support this premise and to confirm the benefits of employing color adaptation techniques with the template matching algorithm follow next.

TABLE 2: Convergence results.

Color transfer implementation	Pt-Pt errors (mean/std)	
	Before fitting	After fitting
None (on grayscale)		16.10/13.76
None (on color)	21.00 / 2.73	14.78/11.48
$I_1I_2I_3$		7.91/4.46
$l\alpha\beta$		9.86/5.33

5. EXPERIMENTS

The experiments have been performed on a randomly chosen subset of 16 images from the database in Figure 1. The images have been semiautomatically annotated and the set of annotations has been used as the ground truth for calculating the boundary errors, which give an objective measure for the fitting quality of the face model. The boundary errors are measured between the exact shape in the image frame (obtained from the ground truth annotations) and the optimized model shape in the image frame. The boundary error is calculated as the *point-to-point* (*Pt-Pt*) error, which is given by the Euclidian distance between the two shape vectors of concatenated x and y coordinates of the landmark points. The mean and standard deviation of *Pt-Pt* errors is used to evaluate the boundary errors over a whole set of images. The results are summarized in Table 2.

An implementation based only on the intensity (gray scale) component has also been tested. The gray scale images have been obtained by applying the standard mix of *RGB* components in (22):

$$Grayscale = 0.30R + 0.59G + 0.11B. \qquad (22)$$

The initial results (no color adaptation) show a slight gain in the fitting accuracy over the gray scale implementation when color information is added. However, significant increase in face segmentation accuracy can be observed when adapting the color of the texture template using color transfer techniques. It can also be noted that the implementation based on $I_1I_2I_3$ color space performs slightly better in terms of segmentation accuracy, although subjectively better color adaptation results have been observed when using the $l\alpha\beta$ color space. This can be explained by the fact that the $I_1I_2I_3$ color space representation is more suitable to be used together with the fitting algorithm which is implemented in the *RGB* color space.

The robustness to changes in the illumination conditions was also tested using the Oulu face image database [29]. An example of color adaptation of the texture template for this database is shown in Figure 5.

6. DISCUSSION AND CONCLUSIONS

We analyzed in this paper the possibility of enhancing a face segmentation/tracking method based on texture template matching by means of color image alignment. We also presented a model parameters optimization approach which
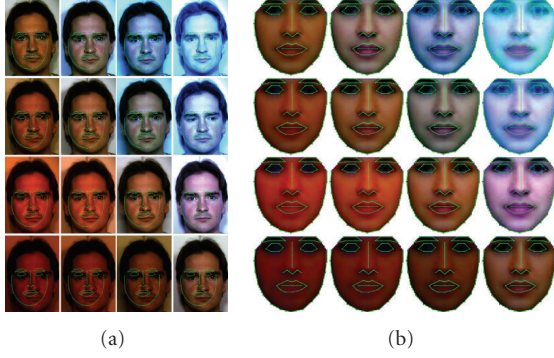
(a)   (b)

FIGURE 5: Example of color adaptation of the texture template under 16 different camera calibration settings and illumination conditions from *Oulu* database (all 4×4 combinations of *horizon*, *incandescent*, *fluorescent*, and *daylight* illuminants). (a) Initial images; (b) color-adapted templates.

minimized the error between the texture template and the warped image texture across the current shape. We employed here the TPS-based warping method which is more robust for head pose variations.

The color alignment techniques make use of the decorrelated color statistics of the current image and template image. Improvements of the accuracy of the segmentation have been demonstrated.

From our experiments, we can conclude that the color-adaptation method for the texture template can also be useful in face tracking applications which employ face modeling techniques similar to the one described in Section 3. In particular, it was shown significant improvements and increased robustness for the case of tracking a face under changes in the illumination conditions, like the change of the type of illuminant. This may be a real change of the illuminant or it could be caused by some wrong white balance setting of the image acquisition device.

APPENDIX

IMAGE WARPING: PRINCIPAL WARPS

The thin plate splines (TPSs)-based warping method, also named principal warps, was first introduced in [23]. It represents a nonrigid registration method, built upon an analogy with a theory in mechanics. Namely, the analogy is made with minimizing the bending energy of a thin metal plate on which pressure is exerted using some point constraints. The bending energy is then given by a quadratic form; the spline is represented as a linear combination (superposition) of eigenvectors of the bending energy matrix:

$$f(x, y) = a_1 + a_x x + a_y y + \sum_{i=1}^{p} w_i U(\| (x_i, y_i) - (x, y) \|),$$

$$\text{(A.1)}$$

where $U(r) = r^2 \log(r)$; (x_i, y_i) are the initial control points. $a = (a_1, a_x, a_y)$ defines the affine part, while w defines the nonlinear part of the deformation.

The total bending energy is expressed as

$$I_f = \iint_{\mathbf{R}^2} \left(\left(\frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 f}{\partial y^2} \right)^2 \right) dx \, dy. \quad \text{(A.2)}$$

The surface is deformed such that to have minimum bending energy. The conditions that need to be met so that (A.1) is valid (so that $f(x, y)$ has second-order derivatives) are given by

$$\sum_{i=1}^{p} w_i = 0,$$

$$\sum_{i=1}^{p} w_i x_i = 0; \qquad \sum_{i=1}^{p} w_i y_i = 0. \quad \text{(A.3)}$$

Adding to this the interpolation conditions $f(x_i, y_i) = v_i$, (A.1) can now be written as the linear system in (A.4):

$$\begin{bmatrix} K & P \\ P^T & O \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} v \\ o \end{bmatrix}, \quad \text{(A.4)}$$

where $K_{ij} = U(\|(x_i, y_i) - (x_j, y_j)\|)$, O is a 3×3 matrix of zeros, o is a 3×1 vector of zeros, $P_{ij} = (1, x_i, y_i)$; w and v are the column vectors formed by w_i and v_i, respectively, while $a = [a_1 \, a_x \, a_y]^T$.

REFERENCES

[1] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," in *Proceedings of the 5th European Conference on Computer Vision (ECCV '98)*, pp. 484–498, Freiburg, Germany, June 1998.

[2] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "A comparative evaluation of active appearance model algorithms," in *Proceedings of the 9th British Machine Vison Conference (BMVC '98)*, vol. 2, pp. 680–689, British Machine Vision Association, Southampton, UK, September 1998.

[3] X. Hou, S. Z. Li, H. Zhang, and Q. Cheng, "Direct appearance models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '01)*, vol. 1, pp. 828–833, Kauai, Hawaii, USA, December 2001.

[4] R. Donner, M. Reiter, G. Langs, P. Peloschek, and H. Bischof, "Fast active appearance model search using canonical correlation analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1690–1694, 2006.

[5] J. Matthews and S. Baker, "Active appearance models revisited," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 135–164, 2004.

[6] M. B. Stegmann and R. Larsen, "Multi-band modelling of appearance," *Image and Vision Computing*, vol. 21, no. 1, pp. 61–67, 2003.

[7] G. J. Edwards, T. F. Cootes, and C. J. Taylor, "Advances in active appearance models," in *Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV '99)*, vol. 1, pp. 137–142, Kerkyra, Greece, September 1999.

[8] R. K. Kouassi, J.-C. Devaux, P. Gouton, and M. Paindavoine, "Application of the Karhunen-Loeve transform for natural color images analysis," in *Proceedings of the 31st Asilomar Conference on Signals, Systems & Computers (ACSSC '97)*, vol. 2, pp. 1740–1744, Pacific Grove, Calif, USA, November 1997.

[9] J.-C. Devaux, P. Gouton, and F. Truchetet, "Aerial color image segmentation by Karhunen-Loeve transform," in *Proceedings of the 15th International Conference on Pattern Recognition (ICPR '00)*, vol. 1, pp. 309–312, Barcelona, Spain, September 2000.

[10] J.-C. Devaux, P. Gouton, and F. Truchetet, "Application of the Karhunen-Loeve transform to aerial color image segmentation," in *Proceedings of the 4th International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies (KES '00)*, vol. 1, pp. 373–376, Brighton, UK, August 2000.

[11] J.-C. Devaux, P. Gouton, and F. Truchetet, "Karhunen-Loeve transform applied to region-based segmentation of color aerial images," *Optical Engineering*, vol. 40, no. 7, pp. 1302–1308, 2001.

[12] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103–108, 1990.

[13] Y. Ohta, T. Kanade, and T. Sakai, "Color information for region segmentation," *Computer Graphics and Image Processing*, vol. 13, no. 3, pp. 222–241, 1980.

[14] J. Mukherjee, "MRF clustering for segmentation of color images," *Pattern Recognition Letters*, vol. 23, no. 8, pp. 917–929, 2002.

[15] P. Mohapatra, P. Nanda, and S. Panda, "Color image segmentation using MRF model and simulated annealing," in *Proceedings of Soft Computing Technique for Engineering Applications (SCT '06)*, NIT, Rourkela, India, March 2006.

[16] D. L. Ruderman, T. W. Cronin, and C.-C. Chiao, "Statistics of cone responses to natural images: implications for visual coding," *Journal of the Optical Society of America A*, vol. 15, no. 8, pp. 2036–2045, 1998.

[17] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer Graphics and Applications*, vol. 21, no. 5, pp. 34–41, 2001.

[18] A. Maslennikova and V. Vezhnevets, "Interactive local transfer between images," in *Proceedings of the International Conference on Computer Graphics & Vision (GraphiCon '07)*, pp. 75–78, Moscow, Russia, June 2007.

[19] J.-H. Lee, B.-H. Chang, and S.-D. Kim, "Comparison of colour transformations for image segmentation," *Electronics Letters*, vol. 30, no. 20, pp. 1660–1661, 1994.

[20] C. Goodall, "Procrustes methods in the statistical analysis of shape," *Journal of the Royal Statistical Society B*, vol. 53, no. 2, pp. 285–339, 1991.

[21] J. J. Gerbrands, "On the relationships between SVD, KLT and PCA," *Pattern Recognition*, vol. 14, no. 1–6, pp. 375–381, 1981.

[22] G. J. Edwards, C. J. Taylor, and T. F. Cootes, "Interpreting face images using active appearance models," in *Proceedings of the 3rd IEEE International Conference on Face & Gesture Recognition (FG '98)*, pp. 300–305, Nara, Japan, April 1998.

[23] F. L. Bookstein, "Principal warps: thin-plate splines and the decomposition of deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 6, pp. 567–585, 1989.

[24] G. Donato and S. Belongie, "Approximate thin plate spline mappings," in *Proceedings of the 7th European Conference on Computer Vision (ECCV '02)*, pp. 21–31, Copenhagen, Denmark, May 2002.

[25] R. K. Beatson and W. A. Light, "Fast evaluation of radial basis functions: methods for two-dimensional polyharmonic splines," *IMA Journal of Numerical Analysis*, vol. 17, no. 3, pp. 343–372, 1997.

[26] A. Zandifar, S.-N. Lim, R. Duraiswami, N. A. Gumerov, and L. S. Davis, "Multi-level fast multipole method for thin plate spline evaluation," in *Proceedings of the International Conference on Image Processing (ICIP '04)*, vol. 3, pp. 1683–1686, Singapore, October 2004.

[27] P. A. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[28] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," in *Proceedings of the 2nd European Conference on Computational Learning Theory (EuroCOLT '95)*, vol. 904, pp. 23–37, Barcelona, Spain, March 1995.

[29] M. Soriano, E. Marszalec, and M. Pietikäinen, "Color correction of face images under different illuminants by RGB eigenfaces," in *Proceedings of the 2nd International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA '99)*, pp. 148–153, Washington DC, USA, March 1999.