A Skin Tone Detection Algorithm for an Adaptive Approach to Steganography

Abbas Cheddad, Joan Condell, Kevin Curran and Paul Mc Kevitt
School of Computing and Intelligent Systems, Faculty of Computing and Engineering
University of Ulster, BT48 7JL, Londonderry, Northern Ireland, United Kingdom
Email: cheddad-a@email.ulster.ac.uk

Abstract. Challenges face biometrics researchers and particularly those who are dealing with skin tone detection include choosing a colour space, generating the skin model and processing the obtained regions to fit applications. The majority of existing methods have in common the de-correlation of luminance from the considered colour channels. Luminance is underestimated since it is seen as the least contributing colour component to skin colour detection. This work questions this claim by showing that luminance can be useful in the segregation of skin and non-skin clusters. To this end, here we use a new colour space which contains error signals derived from differentiating the grayscale map and the non-red encoded grayscale version. The advantages of the approach are the reduction of space dimensionality from 3D, RGB, to 1D space advocating its unfussiness and the construction of a rapid classifier necessary for real time applications. The proposed method generates a 1D space map without prior knowledge of the host image. A comprehensive experimental test was conducted and initial results are presented. This paper also discusses an application of the method to image steganography where it is used to orient the embedding process since skin information is deemed to be psycho-visually redundant.

Keywords: Luminance, colour transform, skin tone detection, steganography, object oriented embedding.

1. Introduction

Detecting human skin tone is of utmost importance in numerous applications such as, video surveillance, face and gesture recognition, human computer interaction, human pose modelling, image and video indexing and retrieval, image editing, vehicle drivers' drowsiness detection, controlling users' browsing behaviour (e.g., surfing indecent sites) and steganography. Detection of human skin tone is regarded as a two-class classification problem, and has received considerable attention from researchers in recent years [1, 2], especially those who deal with biometrics and computer vision aspects.

Modelling skin colour implies the identification of a suitable colour space and the careful setting of rules for cropping clusters associated with skin colour. Unfortunately, most approaches to date tend to put the illumination channel in the "non useful" zone and therefore act instead on colour transformation spaces that de-correlate luminance and chrominance components from an *RGB* image. It is important to note that illumination and luminance are defined slightly differently as they depend on each other. As this may cause confusion, for simplicity, here we will refer to both of them as the function of response to incident light flux or the brightness. Abadpour and Kasaei [3] concluded that "in the *YUV*, *YIQ*, and YC_bC_r colour spaces, removing the illumination related component (Y) increases the performance of skin detection process". Others [4, 5] were in favour of dropping luminance prior to any processing as they were convinced that the mixing of chrominance and luminance data makes *RGB* basis marred and not a very favourable choice for colour analysis and colour based recognition. Therefore, luminance and chrominance are always difficult to tease apart unless the *RGB* components are transformed into other colour spaces, and even then these spaces do not guarantee total control over luminance. Comprehensive work exists which discusses in depth different colour spaces and their performance [3, 6, 7].

Albiol *et al.* [8] and Hsieh *et al*. [9] show that choosing colour space has no implication on the detection given an optimum skin detector is used; in other words all colour spaces perform the same. Analogous to this, Phung *et al*. [7] show that skin segmentation based on colour pixel classification is largely unaffected by the choice of the colour space. However, segmentation performance degrades when only chrominance channels are used in classification. In chrominance based methods, some valuable skin colour information will be lost whilst attempting to separate luminance from chrominance according to [10]. Shin *et al.* [11] question the benefit of colour transformation for skin tone detection, e.g., *RGB* and non-*RGB* colour spaces. Jayaram *et al.* [12] conclude that the illumination component provides different levels of information on the separation of skin and non-skin colour, and thus absence of illumination does not help boost performance. This significant conclusion was drawn based on experiments on different colour transformations with and without illumination inclusion. Their data set comprises 850 images. Among those who incorporate illumination are Lee and Lee [13], where they cluster human skin tone in the 3D space of YC_bC_r transformation.

The proposed method goes a step further and shows that the abandoned luminance component carries considerable information on skin tone; the experiments herein lend some support to this hypothesis. Many colour spaces used for skin detection are simply linear transforms from *RGB* and as such share all the shortcomings of *RGB*[1].

Probability based classifiers are also developed to segregate skin tone regions such as the *Bayes* classifier used in [14]. Additionally, these authors take advantage of inter-frame dependencies in video files. At first, the histogram of the skin pixels and non-skin pixels of the present frame is determined; then the conditional probability of each pixel belonging to the skin area and non-skin area is computed respectively. Next the ratio of these two conditional probabilities is computed. Finally, this ratio is compared with a threshold to determine its property as skin pixel or non-skin pixel.

The following sections are organised as follows: Section 2 discusses related work, Section 3 sets out the proposed method followed by experimental results in Section 4. Section 5 examines the application of the method to steganography and finally conclusions with future work are given in Section 6.

2. Related Work: Human Skin Modelling

Colour transformations are of paramount importance in computer vision. There exist several colour spaces including: *RGB, CMY, XYZ, xyY, UVW, LSLM, L*a*b*, L*u*v*, LHC, LHS, HSV, HSI, YUV, YIQ*, YC_bC_r [15]. The native representation of colour images is the *RGB* colour[2] space which describes the world view in three colour matrices: *Red* (*R*), *Green* (*G*) and *Blue* (*B*). Luminance is present in this space and thus various transforms are intended to extract it out.

2.1 Orthogonal Colour Space (YC_bC_r)

The Y, C_b and C_r components refer to Luminance, Chromatic blue and Chromatic red respectively. This is a transformation that belongs to the family of television transmission colour spaces. This colour space is used extensively in video coding and compression, e.g., MPEG, and is perceptually uniform [16]. Moreover, it provides an excellent space for luminance and chrominance separability [17]. Y is an additive combination of R, G and B components and hence preserves the high frequency image contents; the subtraction of Y in Eq. 1 cancels out the high frequency (Y) [18]. Given the triplet *RGB*, the YC_bC_r transformation can be calculated using the following system[3]:

1 See Colour Space Conversions, <www.poynton.com/PDFs/coloureq.pdf>.
2 Couleur, <http://www.couleur.org/index.php>, dedicated to colour analysis with tools for multi-spectral image analysis.
3 Note: the transformation formula for this colour space depends on the used recommendation.

$$YC_bC_r : \begin{cases} Y = 0.299R + 0.587G + 0.114B \\ C_b = 0.56(B - Y) \\ C_r = 0.71(R - Y) \end{cases} \quad (1)$$

Hsu *et al*. [4] used C_bC_r for face detection in colour images. They developed a model where they noticed a concentration of human skin colour in C_bC_r space. These two components were calculated after performing a lighting compensation that used a "reference white" to normalise the colour appearance. They claimed that their algorithm detected fewer non face pixels and more skin-tone facial pixels. Unfortunately, the testing experiments that were carried out using their algorithm were not in reasonable agreement with this assertion. Some of these results are reported here. Figure 1 describes the algorithm. Similarly, Yun *et al.* [19] used Hsu's algorithm with an extra morphological step where they propose a colour based face detection algorithm in the YC_bC_r color space. The use of the illumination compensation method and a morphology closing was to overcome the difficulty of face detection applicable to video summary. Shin *et al.* [11] showed that the use of such colour space gives better skin detection results compared to seven other colour transformations. The eight colours studied are: *NRGB* (normalized *RGB*), *CIEXYZ*, *CIELAB*, *HSI*, *SCT* (Spherical Coordinate Transform), YC_bC_r, *YIQ*, and *YUV*. *RGB* was used as a baseline performance. For each colour space they dropped its illumination component to form 2D colour.

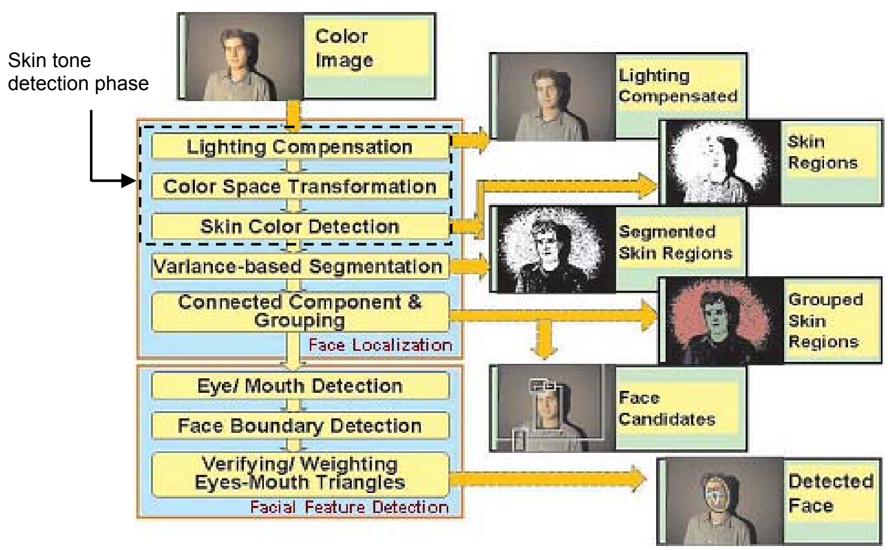

Fig. 1. The system provided by Hsu *et al*. [4].

Choudhury *et al*. [20] developed a method tailored to fit, *File Hound,* which is a field analysis software used by law enforcement agencies during their forensic investigations to harvest any pornographic images from a hard drive. They propose a hybrid algorithm where the compound *RGB* and YC_bC_r based methods are exploited. They notice that the *RGB* method's disadvantage is compensated in YC_bC_r and vice versa. To this end, only relatively large regions which have been missed by the *RGB* filter are re-filtered through a YC_bC_r filter. Time complexity of their approach was not discussed.

Zhao *et al.* [21] construct a vector comprising a blend of different selected components from different colour spaces of which *Cr* was present. *Principal component analysis* (PCA) was applied to this feature vector to find the main orthonormal axes which maximally de-correlate the sample data. A *Mumford-Shah* model was used to segment the image. All of the regions were then traversed to calculate ratio of skin pixels to total pixels within each individual region. Only those regions whose ratio reaches the statistical value would then be regarded as skin regions. Their method entails off-line training and therefore its generalization is questioned.

Hsu *et al*. [4] algorithm was chosen by Shaik and Asari [22] to track faces of multiple people moving in a scene using *Kalman* filters. Zhang and Shi [23] took the same approach with some modifications when the brightness of the face in an image was low. Their method is almost identical to [4] except that they pre-process the image by setting all pixels below 80 to zero in all three primary colours (i.e., *RGB*). They claim their method works better under low brightness mainly due to the pre-processing phase. In order to avoid the extra computation required in conversion from *RGB* to *HSV*, Wong *et al.* [24] use the YC_bC_r colour model and developed a metric that utilises all the components namely Y, C_b and C_r.

2.2 Log Opponent and HSV

The human visual system incorporates colour-opponency and so there is a strong perceptual relevance in this colour space [25]. The Log-Opponent (LO) uses the base 10 logarithm to convert *RGB*[4] matrices into I, R_g, B_y as shown in Eq. 2:

$$\mathrm{IR_gB_y}:\begin{cases}\mathrm{I}=\mathrm{L(G)}\\ \mathrm{R_g}=\mathrm{L(R)}-\mathrm{L(G)}\\ \mathrm{B_y}=\mathrm{L(B)}-(\mathrm{L(R)}+\mathrm{L(G)})/2\end{cases} \tag{2}$$
$$\text{where, } \mathrm{L(x)}=105+\log_{10}(\mathrm{x}+1).$$

[4] Note that this work does not assume a particular range for the *RGB* values.

This method uses what is called hybrid colour spaces. The fundamental concept behind hybrid colour spaces is to combine different colour components from different colour spaces to increase the efficiency of colour components to discriminate colour data. Also, the aim is to lessen the rate of correlation dependency between colour components [26]. Here, two spaces are used, namely *IRGBy* and *HS* from the *HSV* (Hue, Saturation and Value) colour space. *HS* can be obtained by applying a non-linear transformation to the *RGB* colour primaries as shown in Eq. 3. A texture amplitude map is used to find regions of low texture information. The algorithm first locates images containing large areas where colour and texture is appropriate for skin, and then segregates those regions with little texture. The texture amplitude map is generated from the matrix *I* by applying 2D median filters. *RGB* to *HSV* transform can be expressed as in Eq.3.

$$\mathrm{HSV}:\begin{cases} \mathrm{H}=\begin{cases}\mathrm{h}, \mathrm{B}\le \mathrm{G}\\ 2\pi-\mathrm{h}, \mathrm{B}>\mathrm{G}\end{cases} \\ \text{where, } \mathrm{h}=\cos^{-1}\dfrac{1/2[(\mathrm{R}-\mathrm{G})+(\mathrm{R}-\mathrm{B})}{\sqrt{(\mathrm{R}-\mathrm{G})^2+(\mathrm{R}-\mathrm{G})(\mathrm{G}-\mathrm{B})}} \\ \mathrm{S}=\dfrac{\max(\mathrm{R},\mathrm{G},\mathrm{B})-\min(\mathrm{R},\mathrm{G},\mathrm{B})}{\max(\mathrm{R},\mathrm{G},\mathrm{B})} \\ \mathrm{V}=\max(\mathrm{R},\mathrm{G},\mathrm{B}) \end{cases} \tag{3}$$

In order to segment potential face regions, Chen *et al*. [27] analyze the colour of the pixels in *RGB* colour space to decrease the effect of illumination changes, and then classify the pixels into face-colour or non-face colour based on their hue, i.e., component *H* in Eq.3. The classification is performed using Bayesian decision rules. Their method degrades when the images contain complex backgrounds or uneven illumination.

2.3 Basic N-rules *RGB* (N*RGB*)

This is a simple yet powerful method to construct a skin classifier directly from the *RGB* composites which sets a number of rules (*N*) for skin colour likelihood. Kovač *et al.* [28] state that *RGB* components must not be close together, e.g., luminance elimination. They utilize the following rules: An R, G, B pixel is classified as skin if and only if:

$$\begin{aligned}&R > 95 \;\&\; G > 40 \;\&\; B > 20\\ &\&\; \max(R, G, B) - \min(R, G, B) > 15\\ &\&\; |R-G| > 15 \;\&\; R > G \;\&\; R > B\end{aligned} \tag{4}$$

Some authors prefer to normalise the *RGB* primaries beforehand. Let the *RGB* denote the normalised colour space, which is expressed in Eq. 5.

$$r = \frac{R}{R+G+B}, g = \frac{G}{R+G+B}, b = \frac{B}{R+G+B} \tag{5}$$

The *b* component has the least representation of skin colour and therefore it is normally omitted in skin segmentation [29].

Abdullah-Al-Wadud and Chae [10] use a colour distance map (CDM) applied to *RGB* colours, although that can be extended to any colour space. They implement an algorithm based on the property of the flow of water to further refine the output using an edge operator. The generated CDM is a grayscale image. The distribution of the distance map is quasi-Gaussian. They also propose an adaptive Standard Skin Colour (SSC) to act as a classifier to vote for skin pixels. The method does not develop any colour space.

2.4 Other Colour Spaces

Porle *et al.* [29] propose a *Haar* wavelet-based skin segmentation method in their aim to address the problem of extracting the arms which are occluded in the torso. The segmentation procedure is performed using six different colour spaces, namely: *RGB, RGB, HSI, TSL, SCT and CIELAB*. They concluded that the *B* component, representing the position between yellow and blue, in the *CIELAB* colour space has the best performance. Obviously, this technique is complex and time consuming as it involves wavelets decomposition.

3. Proposed Skin Tone Detection Method

Illumination is evenly smeared along *RGB* colours in any given colour image. Hence, its effect is scarcely distinguished here. There are different approaches to segregate such illumination. The utilized transformation matrix is defined in Eq.6.

$$\vec{\alpha} = [0.298936021293775390, \quad 0.587043074451121360, \quad 0.140209042551032500]^T \tag{6}$$

where the superscript *T* denotes the transpose operator to allow for matrix multiplication. Let Ψ denote the 3D matrix containing the *RGB* vectors of the host image and let $x \in [1,2,\ldots,n]$, where $n = \text{length}(R) = \text{length}(G) = \text{length}(B)$. Note that this method acts here on the *RGB* colours stored in double precision, i.e., linearly scaled to the interval [0 1]. The initial colour transformation is given in Eq. 7.

$$I(x) = \left(\Psi(r(x), g(x), b(x)) \otimes \vec{\alpha}\right) \tag{7}$$

where $\otimes$ represents matrix multiplication. This reduces the *RGB* colour representation from 3D to 1D space. The vector *I(x)* eliminates the hue and saturation information whilst retaining the luminance. It is therefore regarded formally as a grayscale colour. Next, the algorithm tries to obtain another version of the luminance but this time without taking the *R* vector into account. Most of skin colour tends to cluster in the red channel. The discarding of red colour is deliberate, as in the final stage it will help to calculate the error signal. Therefore, the new vector will have the largest elements taken from *G* or *B*:

$$\hat{I}(x) = \underset{x \in \{1,\dots,n\}}{\arg\max}(G(x), B(x)) \tag{8}$$

Eq. 8 is actually a modification of the way *HSV* (Hue, Saturation and Value) computes the *V* values. The only difference is that the method does not include in this case the red component in the calculation. Then for any value of *x*, the error signal is derived from the calculation of element-wise subtraction of the matrices generated by Eq. 7 and Eq. 8 which can be defined as given in Eq. 9.

$$e(x) = I(x) - \hat{I}(x) \tag{9}$$

Note that $e(x)$ must employ neither truncation nor rounding.

Creating a *skin probability map* (SPM) that uses an explicit threshold based skin cluster classifier which defines the lower and upper boundaries of the skin cluster is crucial to the success of the proposed technique. A collection of 147852 pixel samples was gathered from different skin regions exhibiting a range of races with extreme variation of lighting effect. After transformation using the proposed method, the projection of data admits a distribution that could be easily fit into a Gaussian curve using Expectation Maximization (EM) method which is an approximation of Gaussian Mixture Models (GMM) as shown in Figure 2. It is also clear that there are no other Gaussians hidden in the distribution.

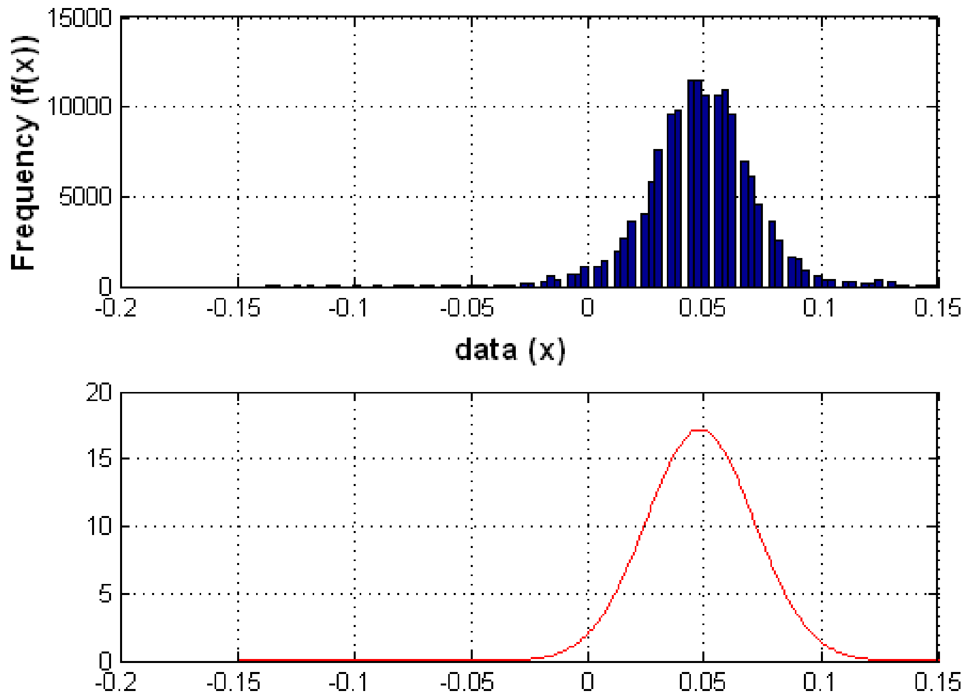

Fig 2. Frequency distribution of the data (top) and its Gaussian curve fit (bottom).

To identify the boundaries, some statistics need to be computed. Let μ and σ denote the mean and standard deviation of the above distribution, and let Δ_{left} and Δ_{right} denote the distances from μ on the left and right hand side respectively. The boundaries are determined based on Eq. 10.

$$\begin{aligned} &\mu - (\Delta_{\text{left}} * \sigma) \approx 0.02511 \\ &\mu + (\Delta_{\text{right}} * \sigma) \approx 0.1177 \end{aligned} \tag{10}$$

Where Δ_{left} and Δ_{right} are chosen to be 1 and 3 sigma away from μ respectively to cover the majority of the area under the curve. Hence, the precise empirical rule set for this work is given in Eq. 11.

$$f_{\text{skin}}(x) = \begin{cases} 1 \text{ if} \quad 0.02511 <= e(x) <= 0.1177 \\ 0 \qquad\qquad \text{otherwise.} \end{cases} \tag{11}$$

This work claims, based on extensive experiments, that this rule pins down the optimum balanced solution. Even though the inclusion of luminance was adopted the 3D projection of the three matrices $I(x), \hat{I}(x), e(x)$ shows clearly that the skin tone clusters around the boundaries given in Eq. 11. This is shown in Figure 3. The dark red dot cloud represents the region where skin colour tends to cluster, i.e., the area bounded by a rectangle. Notice how compact the skin tone is, using the proposed method.

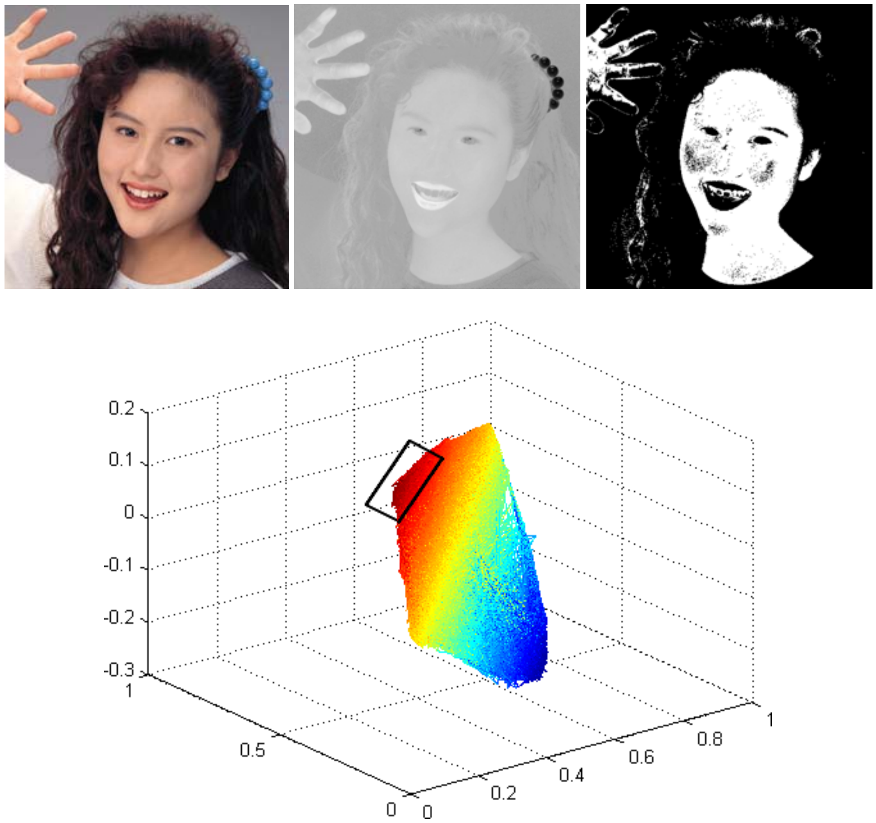

Fig 3. Skin tone segmentation using the proposed method: (top and left to right) original image, result of applying Eq. 9, result of applying Eq. 11, and skin tone cluster in a 3D mesh respectively.

This practical example contradicts the claim reported previously [4] showing the deficiency of using luminance in modelling skin tone colour. The hypothesis that this work wants to support is that luminance inclusion does increase separability of skin and non-skin clusters. In order to provide evidence for this hypothesis, the proposed algorithm was tested on different *RGB* images with different background and foreground complexities. Some images were selected exposing uneven transition in illumination to demonstrate the robustness of the algorithm.

4. Skin Tone Detection: Results and Discussion

For unconvincing reasons, illumination was abandoned by researchers who instead tackled the problem of skin colour detection thinking such a channel had no relevant information for extracting and classifying skin colour pixels. It will be shown that illumination involvement can significantly increase the robustness of the detector. However, like all existing algorithms, it is not yet intelligent enough to discriminate whether a scene posses a skin colour or something that looks similar. The proposed colour model and the classifier can cope with difficult cases encapsulating bad and uneven lighting distribution and shadow interferences. Consequently, these results respond evidently to those authors who arguably questioned the effectiveness of the use of illumination based on its inherent properties. The proposed algorithm outperforms both YC_bC_r and *NRGB* which have attracted many researchers to date. Figure 4 exemplifies how inherent properties of luminance can aid performance if handled intelligently. Notice how the proposed colour space is not affected by the colour distribution which enabled the system to detect skin tone with better efficiency.

Fig 4. Skin detection in an arbitrary image: (left to right) original input image (image 8 in Table 1), skin tone detected by [4], by [25] and by the proposed method in this work respectively.

Figure 5 shows the test images from an Internet database and the corresponding detected skin regions of each algorithm. As shown, the proposed algorithm is insensitive to false alarms. Therefore, it has the least false negative pixels compared to the other three methods, which renders the output cleaner in terms of noise interference. The supreme advantage that the proposed method offers is the reduction of dimensionality from 3D to 1D, which contributed enormously to the algorithm's speed as can be seen in Table 1. These results were obtained using an Intel Pentium Dual Core Processor CPU with Memory Dual-Channel 1024MB (2x512) 533MHz DDR2 SDRAM and 1.6GHz and by using MATLAB Ver. 7.0.1.24704 with IP Toolbox Ver. 5.0.1. It can be seen in Table 1 that the computational time required by some other methods depends on the processed image's content as the processing time is different for images even though they have the same dimensions.

In addition to the arbitrary still images from the Internet, we tested the algorithm on a larger benchmark, i.e., 150 image frames from the popular video "Suzie.avi". This movie sequence is chosen to test for the confusion that hair may cause. Depicted in Figure 6 are some frame samples and the hand labelled ground truth models. Figure 7 shows the graphical performance analysis of the proposal against those reported here. As can be seen, the proposed method is by far the most efficient in that it preserves lower rates for the dual false ratios while securing a high detection rate among all methods (see Figure 7). Figure 8 shows the first four, hand labelled, frames from "Sharpness.wmv" and the overall performance graphed in Figure 9. The video file is used by Windows Media Centre to calibrate the computer monitor by modifying frame sharpness which is suitable for testing the consistency in the performance of the algorithm.

Fig 5. Performance analysis: (left column to right) original images, outputs of [4], [25], [28] and of the proposed method respectively. Shown are some samples from the Internet database that appear in Table 1, where the top corresponds to image 1 and the bottom to image 2.

Table 1. Comparison of computational complexity of the proposed method against other methods [4], [25] and [28] on 12 images obtained from the Internet database of which samples are shown in Fig 5.

Image #	**Number of Pixels**	**Time elapsed in seconds**			
		[4]	**[25]**	**[28]**	**Proposed**
1	840450	0.5160	33.515	7.796	0.125
2	478518	0.4060	22.094	4.156	0.047
3	196608	0.2970	4.547	2.188	0.062
4	196608	0.3280	3.563	1.906	0.062
5	849162	0.5160	33.062	7.531	0.078
6	850545	0.6090	39	8.343	0.062
7	849162	0.6090	39.219	6.641	0.078
8	849162	0.5160	39.172	8.484	0.078
9	849162	0.6100	38.203	6	0.078
10	7750656	3.1720	> 600 *	54.86	0.562
11	982101	0.6410	79.469	7.297	0.078
12	21233664	9.3910	> 600 *	144	1.531

(*) the Log algorithm [25] did not converge for more than 10 min which forced us to halt its process.

Fig 6. The first four frames from a standard testing video sequence: (top) original extracted frames, (bottom) the corresponding Ground Truth from the 150 manually cropped frames.

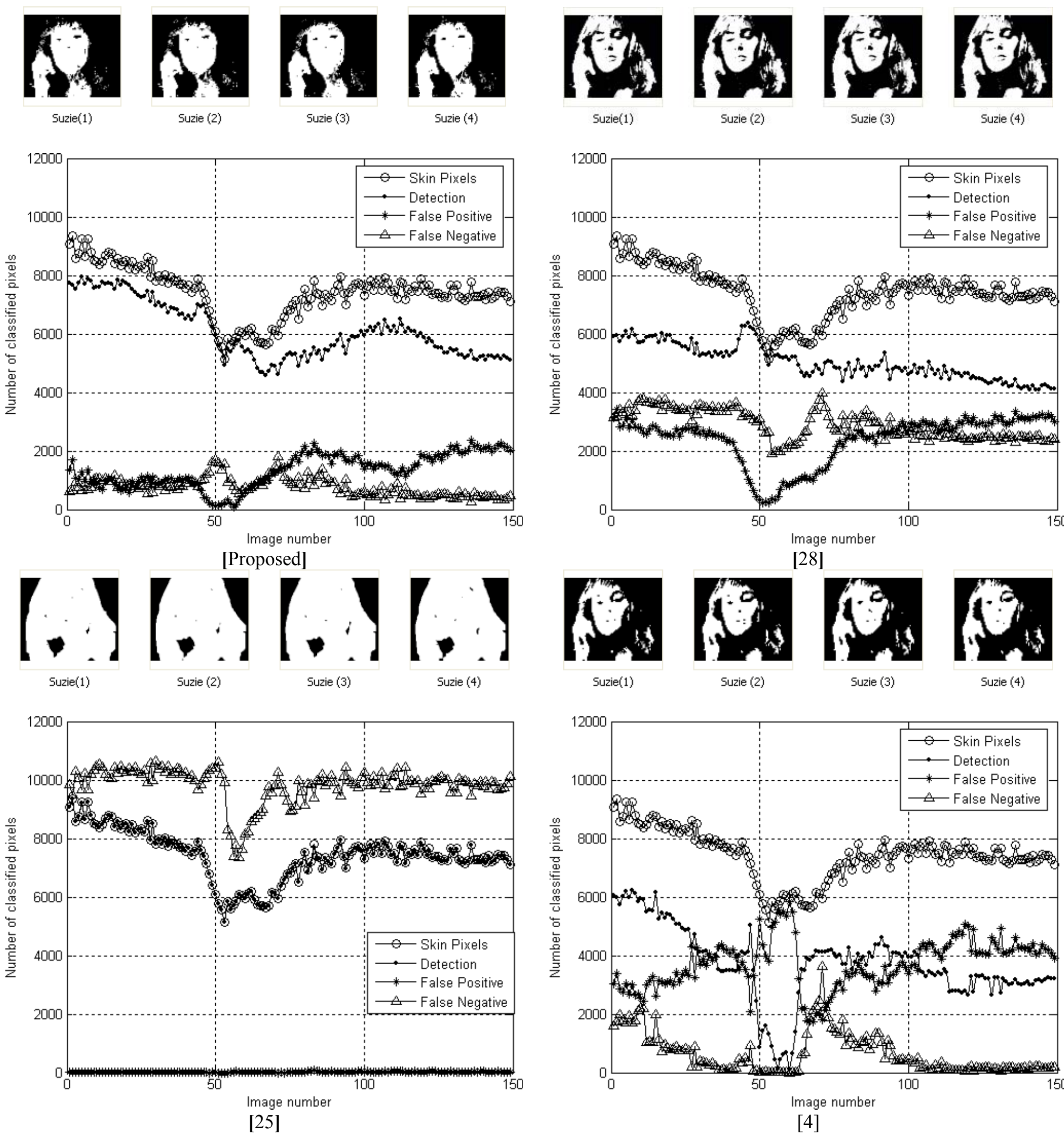

Fig 7. Performance comparison of different methods: the first four frames and performance analysis on the entire 150 frames of "Suzie.avi".

Fig 8. The first four frames from Dell^{TM} video sequence for display testing "Sharpness.wmv": (top) original extracted frames, (bottom) the corresponding Ground Truth.

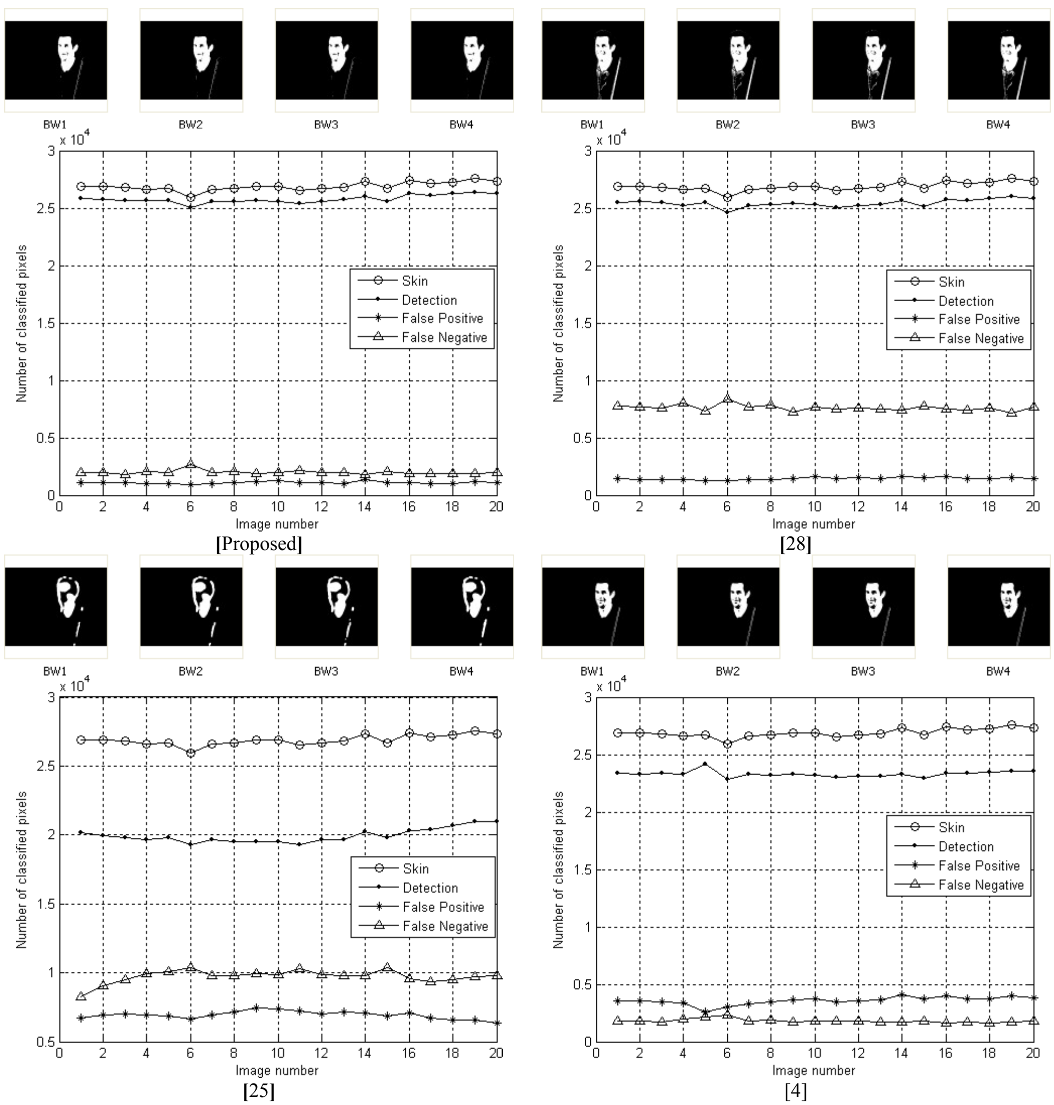

Fig 9. Performance comparison of different methods: the first four frames and performance analysis on the entire 20 frames of "Sharpness.wmv".

Having discussed the contribution for skin tone segmentation and shown the generic experiments, the following section makes explicit the link to the concept of object oriented embedding (OOE).

5. Skin Tone Detection for an Adaptive Approach to Steganography

Steganography is the science of concealing the very existence of data in another transmission medium. It does not replace cryptography but rather boosts the security using its obscurity features. Steganography has various useful applications such as for human rights organizations (as encryption is prohibited in some countries [30]), smart IDs where individuals' details are embedded in their photographs (content authentication) [31], data integrity by embedding checksum [32], medical imaging and secure transmission of medical data [33] to name few. Different algorithms have been proposed to implement steganography in digital images. They can be categorized into three major categories, algorithms in the spatial domain such as *S-Tools* [34], algorithms in the transform domain, e.g., *F5* [35], and algorithms taking an adaptive approach combined with one of the former two methods, e.g., *ABCDE* (A Block-based Complexity Data Embedding) [36]. Most of the existing steganographic methods rely on two factors: (1) the secrecy of the key and (2) the robustness of the steganographic algorithm which can be made public (known as Kerkhoff's principle in cryptography). All of the above tools, along with the majority of other introduced techniques, suffer from intolerance to any kind of geometric distortion applied to the stego-image. For instance, if rotation or translation occurs, all of the hidden data will be lost.

A remedy to this problem could be achieved through incorporating computer vision into the process such as the one formulated previously in Section 3. The concept of object-oriented steganography now becomes one of finding clusters of skin areas in the image 3D space. Recognising and tracking elements in a given carrier while embedding can help survive major image processing attacks and compression. This manifests itself as an adaptive intelligent processing where the embedding process affects only certain *Regions Of Interest* (ROI) rather than the entire image. With developments in Computer Vision (CV) and pattern recognition disciplines, this method can be fully automated and unsupervised. These elements (ROIs), e.g., skin regions, can be adjusted in perfectly undetectable ways. The majority of steganography research to date has overlooked the fact that utilising objects within images can strengthen the embedding robustness - with one exception. Cheddad *et al.* [37] incorporate computer vision to track and segment skin regions for embedding under the assumption that skin tone colour provides better embedding imperceptibility. The algorithm begins by first identifying probable human skin segments as shown in Eq. 12.

$$C = C_{bg} \cup C_{fg}, \quad \text{where, } C_{fg} \in \{\bigcup_{i=1}^{n} S_i\}, S_i \cap S_j = \varnothing, \forall i \neq j \tag{12}$$

C, C_{bg}, and C_{fg} denote the cover image, the background regions and the foreground regions respectively. $\varnothing$ denotes the empty set and (S_1, $S_{2,\ldots,}$ S_n) are connected subsets that correspond to skin regions. Based on experimentation, it is found that embedding into these regions produces less distortion to the carrier image compared to embedding in a sequential order or in any other areas. Such phenomena result from the fact that the eye does not respond with equal weight of sensitivity to all visual information. This is consistent with the claim that certain information simply has less relative importance than other information in the human visual system. This information is said to be psycho-visually redundant since it can be altered without significantly impairing the quality of the image perception [38]. Human presence in digital photography and video files encourages such an approach. In this context, the postulation of the above skin model would definitely help in the case of image translation as it is invariant to such distortions.

With reference to Eq. 13, if the cover image is geometrically transformed by a translation of t_x, along x axis, and t_y, along y axis, in such a way that the new coordinates are given by:

$$C'\begin{bmatrix} x' \\ y' \end{bmatrix} = C\begin{bmatrix} x + t_x \\ y + t_y \end{bmatrix} \tag{13}$$

then each detected skin blob will be transformed likewise with the same distance to the origin as appears in Eq. 14.

$$S_i'\begin{bmatrix} x' \\ y' \end{bmatrix} = S_i\begin{bmatrix} x + t_x \\ y + t_y \end{bmatrix}, \forall i \in \{1,\ldots,n\} \tag{14}$$

Skin regions are extracted based on colour tone; therefore, are undisturbed by translation [39].

To cope with rotation, it is sufficient to locate face features, i.e., eyes, based on the method described in [21]. Let the distance between the two centres of the eyes be D, then the geometrical face model and its relative distances can be described as follows (refer to Figure 10) [40]:

- Centre of the ellipse (x_0, y_0): is the centre of the distance between the two eyes
- Minor axis length (a) : is the distance between the two eyes where both eye centres lie on each side of the ellipse
- Major axis length (b): is $2D$ where D denotes the distance between the two eyes

- Angle ($\angle$): the ellipse must have the same orientation as the detected face. A face orientation can be easily determined based on the angle made by the baseline (a line connecting both eyes) and the horizontal *x* axis

Salient features form reference points that dictate the orientation of embedding and thus aid recovery from rotational distortions (see Figures 11 and 12). Experiments on other types of attack are shown in Figure 13.

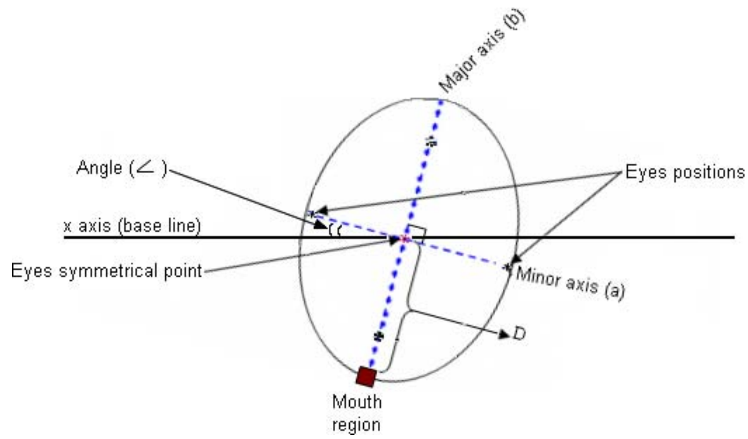

Fig. 10. Elliptical model formed by face features.

Rotation about the origin is defined as in Eq.15.

$$x' = x\cos\theta - y\sin\theta, y' = x\sin\theta + y\cos\theta \tag{15}$$

The angle θ will be determined from the above elliptical model. Hence, if the attacked image is rotated in the opposite direction with the same angle, i.e., $\theta' = -\Delta\theta$ caused by the attack, the method will be able to restore the coordinates and will have the coordinates as shown in Eq. 16.

$$x' = x, y' = y \tag{16}$$

Eq. 16 is used where embedding occurs in the neutralised orientation where $b_{axis} \perp x_{baseline}$. However, the encoder is against 359 options to choose from for the angle as expressed in Eq. 17.

$$\theta' = \theta \pm \alpha\ , \tag{17}$$

where $\alpha \in \{1,2,...,359\}$ denotes an agreed upon scalar which can form another optional secret key. Note that, for simplicity, α here belongs to the discrete space while in practice is continuous. But we encourage the use of discrete values in order to minimise the errors in the recovered bits.

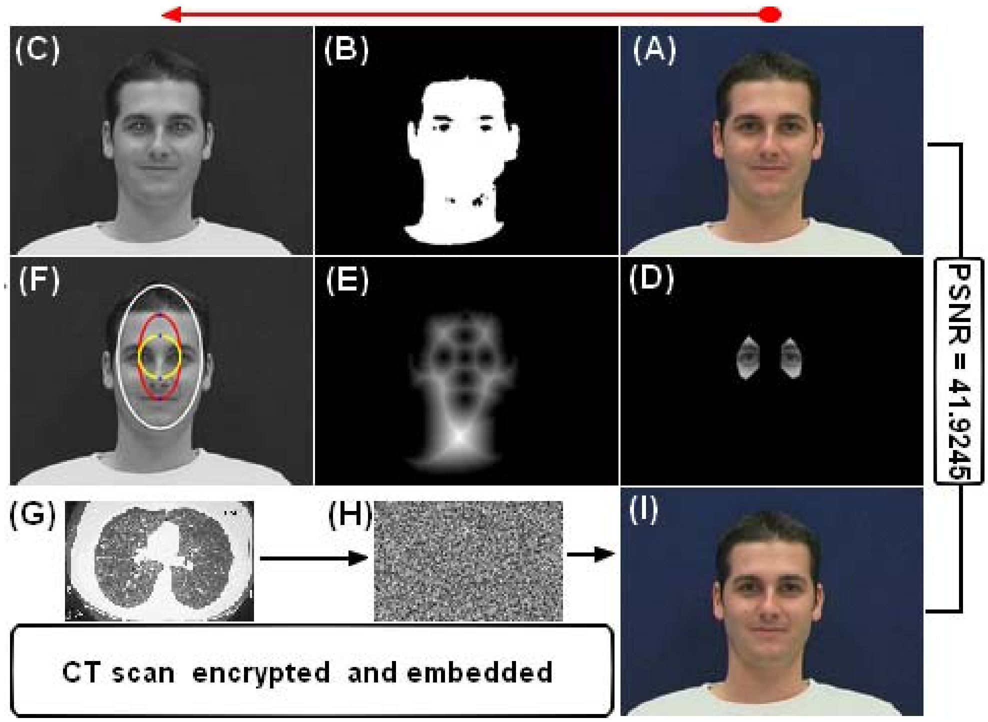

Fig 11. Proposed skin based steganography system concealing medical data in a face image: original image (A), skin blob of the segmented skin area (B), eyes' centroid detection (C), eye regions (D), distance transformation based on face features (E), construction of ellipses [41] (F), CT scan image (G), CT scan encrypted (H) and stego-image carrying the embedded CT image (I). Shown on the right is the PSNR (Peak Signal to Noise Ratio)-measurement for image distortion.

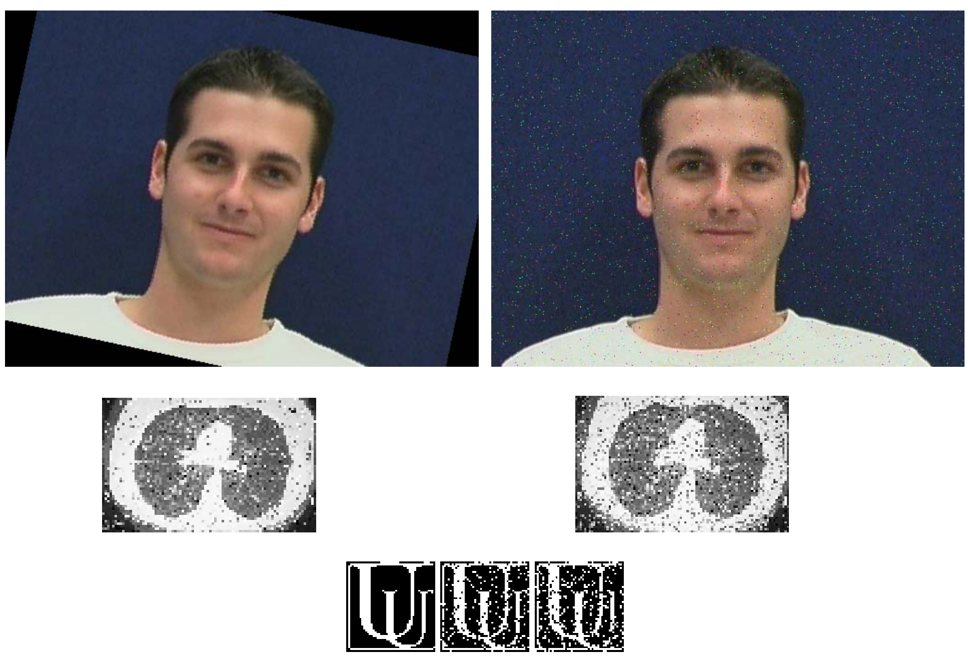

Fig 12. Resistance to image processing attacks: (left) attacked stego-image with a joint attack of cropping and rotation of -12 degrees and the extracted secret data, the little error in the extracted signal is due to interpolation operation (right) attacked with salt and pepper noise and the extracted secret data, (bottom, left to right) successful extraction of embedded data after JPEG compression with quality factors Q=100, Q=80 and Q=75 respectively.

Fig 13. Resistance to other deliberate image processing attacks: (top left) shows the original cover image - *ID01_035.bmp*- obtained from GTAV Face Database[5] along with the image annotation to embed, (top right) is the attacked stego-image and the extracted annotation, (bottom left) attacked stego-image with half transparent frame and the extracted annotation and finally (bottom right) shows an attack on stego-image with translation to the left with an offset=200 pixel and the extracted annotation which is identical to the embedded one.

Typically, targeting specific regions would yield a reduction in space available for embedding, but comes at the benefit of robustness and perception. The embedding takes place in the 1^{st}-level 2D *Haar* DWT (Discrete Wavelet Transform) with the symmetric-padding mode to resist noise impulse and compression. Although algorithms based on DWT experience some losses of data since the reverse transform truncates the values if they go beyond the lower and upper boundaries (i.e., 0-255). Knowing that human skin tone resides along the middle range in the chromatic red of YC_bC_r colour space allows us to embed in the *DWT* of the C_r channel. This would leave the perceptibility of the stego-image virtually unchanged since the changes made in the chrominance will be spread among the *RGB* colours when transformed.

To conclude this section a summary of the drawback of the current steganographic techniques and the main characteristics underlying the proposed method is given. This summary is tabulated in Table 2.

[5] F. Tarrés and A. Rama, GTAV Face Database, http://gps-tsc.upc.es/GTAV/ResearchAreas/UPCFaceDatabase/GTAVFaceDatabase.htm.

Table 2. Drawbacks of current methods and benefits of proposed method.

Method	Descriptions
Spatial domain techniques (i.e., [34])	▪ Large payload but often offset the statistical properties of the image ▪ Not robust against lossy compression and image filters ▪ Not robust against rotation, cropping and translation ▪ Not robust against noise ▪ Many work only on BMP format ▪ Do not address encryption of the payload or use conventional algorithms
DCT based domain techniques (i.e., [35])	▪ Less prone to attacks than the former methods at the expense of capacity ▪ Breach of second order statistics ▪ Breach of DCT coefficients distribution ▪ Work only on JPEG format ▪ Double compressing the file ▪ Not robust against rotation, cropping and translation ▪ Not robust against noise ▪ Modification of quantization table ▪ Do not address encryption of the payload or use conventional algorithms
Proposed	▪ Object oriented ▪ Small embedding space at the benefit of robustness. Resolved by targeting video files which have excellent features for information hiding such as large capacity and good imperceptibility [41] ▪ Resistance to rotation, translation, cropping and moderate noise impulses ▪ No known statistical breach ▪ Resistance to lossy compression ▪ Performs better than DCT algorithms in keeping the carrier distortion to the minimum ▪ Addresses a novel encryption method of the payload [42]

6. Conclusion and Future Work

According to Zhao *et al.* [43], there are two critical issues for colour-based skin detection: (1) what colour space should be selected? and (2) what segmentation method should be used? This paper addresses a novel colour space where human skin clusters can be classified with carefully selected boundaries. The introduced colour space reduces the *RGB* composite from 3D space to purely 1D space reducing the number of image colours which is salient for segmentation and lossy compression of colour visual information [44]. Also, it would play a vital role in content based video coding [45] and content-based image retrieval (CBIR) such as the one introduced in [46]. The test database used consists of randomly collected images from the Internet that cater for different intrinsic and extrinsic characteristics, 150 frames from the *Suzie.avi* movie and the first 20 frames from *Sharpness.wmv* which were hand labelled to generate quantitative measurements.

Additionally, this work sets in context and gives credence to the proposed hypothesis that luminance inclusion does increase separability of skin and non-skin clusters as the shown results agree reasonably with the speculated hypothesis. It is hoped that this work has established that luminance inclusion helps increase separability of skin and non-skin clusters. It is worth pointing out that the proposed method does not rely solely on luminance.

Future work will extend experiments to explore if skin colour detection can be improved in the reduced dimensionality space of wavelets although this will increase the computational burden. Also, comparison between different colour space converters will be targeted. This skin detection technique can be applied to information hiding, specifically steganography, which restrains permanently rotation and translation attacks. The main objective of this work reported here is to establish a new approach for achieving object-oriented steganography.

References

[1] M. Corey, F. Farzam and J.H. Chong, The effect of linearization of range in skin detection, in: Proceedings of IEEE International Conference on Information, Communications and Signal Processing, Singapore, 10-13 December 2007, pp. 1-5.

[2] U.A. Khan, M.I. Cheema and N.M. Sheikh, Adaptive video encoding based on skin tone region detection, in: Proceedings of IEEE Students Conference, Pakistan, 16-17 August 2002, vol (1), pp. 129-34.

[3] A. Abadpour and S. Kasaei, Pixel-based skin detection for pornography filtering, Iranian Journal of Electrical and Electronic Engineering, 1(3)(2005) 21-41.

[4] R.L. Hsu, M. Abdel-Mottaleb and A.K. Jain, Face detection in color images, IEEE Trans. Pattern Analysis and Machine Intelligence, 24(5)(2002) 696-702.

[5] V. Vezhnevets, V. Sazonov and A. Andreeva, A survey on pixel-based skin color detection techniques, in: Proc. Graphicon, Moscow, September 2003, pp. 85-92.

[6] J.B. Martinkauppi, M.N. Soriano and M.H. Laaksonen, Behavior of skin color under varying illumination seen by different cameras at different color spaces, in: Proc. of SPIE, Machine Vision Applications in Industrial Inspection IX, USA, 2001, vol. 4301, pp. 102-113.

[7] S. L. Phung, A. Bouzerdoum and D. Chai, Skin segmentation using color pixel classification: analysis and comparison, IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(1)(2005) 148-154.

[8] A. Albiol, L. Torres, and E.J. Delp, Optimum color spaces for skin detection, in: Proceedings of the IEEE International Conference on Image Processing, Greece, 2001, vol. 1, pp.122-124.

[9] I.S. Hsieh, K.C. Fan and C. Lin, A statistic approach to the detection of human faces in color nature scene, Pattern Recognition, 35(7)(2002) 1583-1596.

[10] M. Abdullah-Al-Wadud and O. Chae, Skin segmentation using color distance map and water-flow property, in: Proceedings of International Conference on Information Assurance and Security, Italy, 8-10 Sept. 2008, pp. 83-88.

[11] M.C. Shin, K.I. Chang and L.V. Tsap, Does colorspace transformation make any difference on skin detection?, in: Proceedings of IEEE Workshop on Applications of Computer Vision, Florida, 3-4 December 2002, pp. 275-279.

[12] S. Jayaram, S. Schmugge, M.C. Shin and L.V. Tsap, Effect of colorspace transformation, the illuminance component, and color modelling on skin detection, in: Proc of IEEE Computer Vision and Pattern Recognition (CVPR'04), Washington, 27th June-2nd July 2004, vol.2, pp. 813-818.

[13] H.J. Lee, and C.C. Lee, Human skin tone detection in YC_bC_r space, Patent No.: US 2005/0207643 A1. Publication date: Sep. 22, 2005. USA.

[14] Y. Liu and C.X. Wang, An improved algorithm of human skin detection in video image based on linear combination of 2-order Markov and Wiener predictor, in: Proceedings of International Symposium on Computer Science and Computational Technology, Shanghai, 20-22 December 2008, vol. 2, pp.665-668.

[15] G. Gomez, On selecting colour components for skin detection, in: Proceedings of International Conference on Pattern Recognition, Quebec, 11-15 Aug 2002, vol.2, pp. 961-964.

[16] M.C. Chi, J.A. Jhu and M.J. Chen, H.263+ region-of-interest video coding with efficient skin-color extraction, in: Proceedings of International Conference on Consumer Electronics, Las Vegas, 7-11 Jan.

2006, pp. 381-382.
[17] M. Beniak, J. Pavlovicova and M. Oravec, Automatic face detection based on chrominance components analysis systems, in: Proceedings of International Conference on Systems, Signals and Image Processing, Slovak Republic, 25-28 June 2008, pp.475-478.
[18] N.X. Lian, V. Zagorodnov and Y.P. Tan, Image denoising using optimal color space projection, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, France, 14-19 May 2006, vol.2, pp. 93-96.
[19] J.U. Yun, H.J. Lee, A.K. Paul and J.H. Baek, Robust face detection for video summary using illumination-compensation and morphological processing, in: Proceedings of IEEE International Conference on Natural Computation, China, 24-27 August 2007, pp. 710-714.
[20] A. Choudhury, M. Rogers, B. Gillam and K. Watson, A novel skin tone detection algorithm for contraband image analysis, in: Proceedings of International Workshop on Systematic Approaches to Digital Forensic Engineering, California, 22 May 2008, pp.3-9.
[21] Y.J. Zhao, S.L. Dai and X. Xi, A Mumford-Shah level-set approach for skin segmentation using a new color space system, in: Proceedings of International Conference on Simulation and Scientific Computing, China, 10-12 October 2008, pp. 307-310.
[22] Z. Shaik, and V. Asari, A robust method for multiple face tracking using Kalman filter, in: Proceedings of IEEE Applied Imagery Pattern Recognition Workshop, Washington, 10-12 October 2007, pp.125-130.
[23] Z. Zhang and Y. Shi, Face detection method based on a new nonlinear transformation of color spaces, in: Proceedings of International Conference on Fuzzy Systems and Knowledge Discovery, China, 18-20 October 2008, vol. 4, pp. 34-38.
[24] K.W. Wong, K.M. Lam and W.C. Siu, A robust scheme for live detection of human faces in color images, Signal Processing: Image Communication, 18(2)(2003) 103-114.
[25] J. Berens and G.D. Finlayson, Log-opponent chromaticity coding of colour space, in: Proceedings of IEEE International Conference on Pattern Recognition, Barcelona, 2000, vol.1, pp. 206-211.
[26] D. Forsyth and M. Fleck, Automatic detection of human nudes, International Journal of Computer Vision, 32(1)(1999) 63-77.
[27] H.Y. Chen, C.L. Huang and C.M. Fu, Hybrid-boost learning for multi-pose face detection and facial expression recognition, Pattern Recognition, 41(3)(2008) 1173-1185.
[28] J. Kovač, P. Peer and F. Solina, Human skin colour clustering for face detection, in: Proceedings of International Conference on Computer as a Tool, Slovenia, 22-24 September 2003, vol.2, pp. 144-148.
[29] R.R. Porle, A. Chekima, F. Wong and G. Sainarayanan, Wavelet-based skin segmentation for detecting occluded arms in human body pose modelling system, in: Proceedings of International Conference on Intelligent and Advanced Systems, Malaysia, 25-28 November 2007, pp.764 -769.
[30] Cryptology and Circumvention, 2003, http://www.frontlinedefenders.org/manual/en/esecman/chapter3_4.html.
[31] A. K. Jain and U. Uludag, Hiding fingerprint minutiae in images, in: Proceedings of Workshop on Automatic Identification Advanced Technologies, New York, 2002, pp.97-102.
[32] C.C. Chang, Y.S. Hu and T.C. Lu, A watermarking-based image ownership and tampering authentication scheme, Pattern Recognition Letters, 27(5)(2006) 439-446.
[33] D. C. Lou, M.C. Hu and J.L. Liu, Multiple layer data hiding scheme for medical images. Computer Standards and Interfaces, 31(2)(2009) 329-335.
[34] A. Brown, 1996, S-Tools, <http://www.jjtc.com/Security/stegtools.htm>.
[35] A. Westfeld, 2001, F5, <http://wwwrn.inf.tu-dresden.de/~westfeld/f5.html>.
[36] H. Hioki, A data embedding method using BPCS principle with new complexity measures, in: Proceedings of Pacific Rim Workshop on Digital Steganography, Japan, July 2002, pp.30-47.
[37] A. Cheddad, J. Condell, K. Curran and P. Mc Kevitt, Biometric inspired digital image steganography, in: Proceedings of the 15th Annual IEEE International Conference and Workshops on the Engineering of Computer-Based Systems (ECBS'08), Belfast, 2008, pp. 159-168.
[38] R. C. Gonzalez and R.E. Woods, Digital image processing, Prentice Hall, 2002, Chapter 8, pp: 417.
[39] A. Cheddad, J. Condell, K. Curran and P. Mc Kevitt, Skin tone based steganography in video files exploiting the YC_bC_r colour space, in: Proceedings of IEEE International Conference on Multimedia and Expo, June 23-26 2008, Hannover, Germany, pp.905-909.
[40] A. Cheddad, D. Mohamad and A. Abd Manaf, Exploiting Voronoi diagram properties in face segmentation and features extraction, Pattern Recognition, 41(12)(2008) 3842-3859.
[41] Z. Zhao, N. Yu, and X. Li, A novel video watermarking scheme in compression domain based on fast motion estimation, in: Proceedings of IEEE International Conference on Communications, China, vol. 2, 2003, pp.1878-1882.
[42] A. Cheddad, J. Condell, K. Curran and P. Mc Kevitt, Securing information content using new encryption method and steganography, in: Proceedings of IEEE International Conference on Digital In formation Management, 2008, London, pp. 563-568.
[43] X. Zhao, F. Boussaid and A. Bermak, Characterization of a 0.18 μ m CMOS color processing scheme for skin detection, IEEE Sensors Journal, 7(11)(2007)1471-1474.
[44] X. Li, T. Yuan, N. Yu, and Y. Yuan, Adaptive color quantization based on perceptive edge protection, Pattern Recognition Letters, 24(16)(2003)3165-3176.

[45] D. Xu, X. Li, Z. Liu and Y. Yuan, Cast shadow detection in video segmentation, Pattern Recognition Letters, 26(1)(2005)91-99.
[46] X. Li, Image retrieval based on perceptive weighted color blocks, Pattern Recognition Letters, 24(12)(2003)1935-1941.