Abstract:

Title: METHOD FOR DETECTION OF OBJECTIONABLE CONTENTS IN STILL OR ANIMATED DIGITAL IMAGES

A method for detecting objectionable contents in a bidimensional pixel image comprises: implementing a first process for detecting geometric features of the image, implementing a second process for detecting skin colors in the image, implementing a third process for detecting object shapes indicative of sexual contents in the image.
METHOD FOR DETECTION OF OBJECTIONABLE CONTENTS IN STILL OR ANIMATED DIGITAL IMAGES

The present invention generally relates to computer-implemented methods for detecting undesirable contents, such as illegal contents or contents intended to a specific public, such as explicit sexual contents for adults.

Background of the invention

With the development of the Internet and its access by people in all categories of ages, cultures, etc, there is a growing need for automatically detecting and filtering out pornographic type contents.

Current methods for detecting and prohibiting access to such types of content, as well as content with racist, violent or other features, are based on detecting text in the contents’ sources.

There is currently no practical and efficient way to detect whether a still image or a video sequence accessible through the Internet is an undesirable content, which should be filtered out.

There have been some efforts to perform such detections, but they have failed to be effective.

The present invention aims at providing a new and effective detection method based on a combination of processes or agents.

Summary of the invention

Accordingly, the present invention provides a method for detecting objectionable content in a bi-dimensional pixel image, comprising:

- implementing a first process for detecting geometric features of the image,
- implementing a second process for detecting skin colors in the image,
- implementing a third process for detecting object shapes indicative of sexual contents in the image.
Preferred but non-limiting aspects of this method are as follows:

- said bi-dimensional pixel image belongs to a video sequence, and the method further comprises:
  - implementing a fourth process for detecting recurrent movement in a succession of images.

- said first process includes a sub-process for detecting circular patterns, said sub-process comprising the following steps:
  - scanning the image with a region-of-interest thumbnail having a size of X x Y pixels substantially smaller than the starting image, and
  - for each of a plurality of individual thumbnails:
    - calculating a linear correlation coefficient between said thumbnail and the same thumbnail rotated through 90° relative to said thumbnail, and
    - identifying said thumbnail as containing a circular pattern if the calculated linear correlation coefficient exceeds a threshold.

- said first process includes a sub-process for detecting areas having axial symmetry in a bi-dimensional pixel image, said sub-process comprising the following steps:
  - scanning the image with a region-of-interest thumbnail having a size of X x Y pixels substantially smaller than the starting image, and
  - for each of a plurality of individual thumbnails:
    - calculating a linear correlation coefficient between said thumbnail and the same thumbnail rotated through 180° relative to said thumbnail, and
    - identifying said thumbnail as containing a circular pattern if the calculated linear correlation coefficient exceeds a threshold.

- X = Y.

- X and Y have values between 16 and 64.

- the first process is performed for all possible thumbnail positions in the image.
* each pixel of the image has values in a color space having N dimensions (such as RGB), and said second process comprises the following steps:
  - scanning the image with a region-of-interest thumbnail having a size of X x Y pixels substantially smaller than the starting image, and
  - for each of a plurality of selected individual thumbnails:
    • in each color subspace having a dimension M lower than N (RG, RB, GB), measuring the distortion of the cloud of two dimensional points corresponding to the thumbnail pixel values in this color subspace, so as to provide distortion coefficients forming a vector of colorimetric context,
    • applying said vector to a neural network previously trained with labeled images to generate color classification.
  - the second further comprises, before applying vector to the neural network, the step of eliminating non eligible areas of the image (too green, too blue, too dark, etc.).
    * said elimination step comprises converting the color pixel values (RGB) into HSL (Hue/Saturation/Lightness) pixel values and applying to said HSL values a gauge eliminating non-skin type hue values, as well as dark pixels and black & white pixels.
    * said neural network outputs degree of membership values among three classes, i.e. normal skin, dark skin, and non-skin.
    * the second process comprises a further step of applying said degree of membership values to a table of fuzzy decision rules for determining a score of skin color associated with an index score of reliability.
    * the second process comprises a step of generating an image of skin color scores for a plurality of eligible areas of the starting image, and a step of noise removal and binahzation to retain as skin areas those of adjacent pixels all have the same binary value.
  - said third process comprises the following steps:
- for an image of a sequence, providing a vector \((D_x, D_y)\) of apparent movement relative to a previous image of the sequence in different positions of the image,

- for a plurality of image positions, determining apparent travel signal values \((S_x, S_y)\) at that position, made of series of vectors values for successive images of the sequence in that position,

- searching for potential periodic characteristics the apparent travel signal values by:
  - removing signal components corresponding to slow and/or linear movements of the camera,
  - computing an autocorrelation function \((FAC)\) of said signal values after said components removal,
  - generating a trigonometric function \((CNpz)\) of said autocorrelation function,
  - computing a correlation between said autocorrelation function and said trigonometric function,
  - comparing the correlation value with a threshold so as to identify the signal values \((S_x, S_y)\) as periodic,
  - determining the frequency of the periodicity of the signal values.

* the third process comprises a step of determining the amplitudes of the periodic signal values.

* the determinations are made independently along two perpendicular directions.

* if the frequencies are essentially identical along the two directions, then the amplitudes of the signal values along the two directions are combined in a vector sum.

* the removal step comprises computing a linear regression of \(S(t)\) among \(t\), so as to remove the trend component thereof.

* said fourth process comprises the following steps:
- providing a plurality of decomposition thumbnails having a size of $X \times Y$ diagrammatically representing different object appearances,
- scanning the image to form region-of-interest (ROI) thumbnails of image portions having different sizes,
- computing a linear correlation between said ROI thumbnails and said decomposition thumbnails with different mutual orientations therebetween,
- generating a signature of each ROI thumbnail made from the linear correlation values obtained with each decomposition thumbnail, and
- using a neural network previously trained with labeled images containing said object for discriminating between signatures to deliver a membership function discriminating between object containing images and object-free images.

* said thumbnails are black & white.
* said fourth process comprises post-processing the membership function, said post-processing including dynamics reshaping by repositioning the slope of the membership function.

The present invention also provides a computer system or computing environment including sets of instructions for performing the above method, as well as information media or carriers containing said sets of instructions.

**Brief description of the drawings**

Other aims, features and advantages of the present invention will appear more clearly from the following detailed description of embodiments thereof, made with reference to the appended drawings, in which:

Fig. 1 illustrates the scanning of a bidimensional pixel image with a region of interest,

Fig. 2 illustrates a process for detecting circular patterns,

Fig. 3 illustrates a process for detecting symmetric patterns,

Fig. 4 is a series of curves illustrating the variability of bi-dimensional color vectors, for use designing a skin color detection process,
Fig. 5A illustrates a membership function used in hue detection for this process,
Fig. 5B illustrates a membership function used in saturation detection for said process,
Fig. 6 is a synoptic representation of a pixel eligibility scheme for the skin color detection process,
Fig. 7 is a synoptic representation of the skin color detection process,
Figs. 8A, 8B and 8C illustrate fuzzy logic rules used in the skin color detection process,
Fig. 9 illustrates a pixel region used for de-noising and binarizing pixel values for the skin color detection process,
Figs. 10A and 10B illustrate two types of optical flows for detecting recurrent movement,
Fig. 11 illustrates a trendless signal generation step used in a recurrent movement detection process of the invention,
Fig. 12 illustrates an autocorrelation step used in said process,
Fig. 13 illustrates a linear correlation coefficient generation step used in said process,
Fig. 14 shows a set of model thumbnails used in a penetration detection process,
Fig. 15 illustrates a neural network used for said process,
Figs. 16A and 16B illustrate two fuzzy logic membership functions used in said neural network,
Fig. 17 shows a set of model thumbnails used in a breast nipple detection process,
Fig. 18 illustrates a neural network used for said process, and
Figs. 19A and 19B illustrate two fuzzy logic membership functions used in said neural network.

Detailed description of a preferred embodiment
As a preliminary note, the following description will refer to certain algorithms that have been previously described in the scientific literature. They are not described again in the description and the skilled person will refer to the corresponding publications.

In addition, the mathematical tools (linear correlation, coefficient of linear regression, polynomial approximation, ...) used in the present invention are referenced by their name, without reminder formulas that are readily available in literature.

1) **Edge and corner detection processes**

The present invention may implement conventional processes for edge detection and corner detection, such as the corner detectors of Harris and Moravec (cf. en.wikipedia.org/wiki/Corner_detection).

2) **Circular area detection process**

The system scans each image entirely with a Region of Interest ("ROI") that contains a thumbnail to be analyzed, having a size of e.g. 32 x 32 pixels. This ROI successively assumes all possible locations in the complete picture, as illustrated in Fig. 1.

For each extracted thumbnail, the system computes a "circular area" score that is obtained by calculating a linear correlation coefficient (ranging between -1 and 1) between two images derived from the same thumbnail being analyzed. This coefficient is however assigned the 0 value when it is negative, so as to obtain a circular area score ranging between 0 and 1.

Preferably, the two images are on the one hand, the thumbnail itself, and on the second hand the same thumbnail after a rotation of an angle of 90°.

As a consequence, the shape in the thumbnail which has the highest similarity with the same shape when turned through 90° has the higher score, and such shapes are the most circular ones, as illustrated in Fig. 2.
3) Detection process for zones presenting axial symmetry

Again, each image is entirely scanned with a Region of Interest (ROI) that contains a thumbnail to be analyzed, having a size of e.g. 32 x 32 pixels), which successively assumes all possible locations in the complete picture, as illustrated in Fig. 1.

For each extracted thumbnail, a "symmetric data" score is computed, which is obtained by calculating a linear correlation coefficient (ranging between -1 and 1) between two images derived from the same thumbnail being analyzed. This coefficient is however assigned the 0 value when it is negative, so as to obtain a circular area score ranging between 0 and 1.

Preferably, the two images are on the one hand, the thumbnail itself, and on the second hand the same thumbnail after a rotation of an angle of 180°.

As a consequence, the shape in the thumbnail which has the highest similarity with the same shape when turned through 180° has the higher score, and such shapes are the most symmetric ones, as illustrated in Fig. 3.

The circular areas and axially symmetrical image portions as detected by the above processes are often found in soft-core or hardcore adult programs.

4) Color recognition process: application to skin detection

There has been some published research work on the detection of skin color based on video color components (RGB, HSL, YUV, C.I.E. Luv and C.I.E. Lab).

Apart from the RGB space which is generally not used for recognition of colors in general (except in the case of very weakly saturated colors), all other spaces have three available components: Brightness, Hue, Saturation.

Most known human skin color recognition concepts try to characterize the skin by a pattern in hue value and a template to saturation.

There are some works that combine 3 variables in a function of mixture, as presented in the thesis of Thierry CARRON, 1995, University of
Savoie, "segmentation of color images in the database luminance hue saturation." The same thesis presents approaches of fuzzy logic for the segmentation of colors.

However, none of these methods can detect skin in a reliable manner, and without an unacceptable false detection rate.

According to this aspect of the present invention, a hundred pornographic videos taken at random from the internet have been studied, and it has been perceived that intra class variability (human skin) among several videos is greater than the inter class variability on a single video.

More broadly, it has been experienced by the inventors that the problem of variability touched all hues. For example, they consistently compared images of "blue sky", and found a wide range of hues from one image to another, which inhibits a simple way to recognize "blue sky". All that is described below in relation with the color detection process is applied to the recognition of skin color, but the process applies to any class of colors.

The above examples, culled from a limited number of video sequences, show us that it is impossible to classify in an absolute way "skin" or "no skin."

In general, there is, over a given image, a color contrast between the skin and the decor. But a color that matches the decor on an image may have exactly the same color as the skin on another image.

The factors of the variability of skin color are mainly:
- intrinsic dissimilarity of skin color;
- color variance due to poor lighting and poor white balance.

In pornographic videos, the white balance is generally deficient, and it is this second factor that generates the majority of hue variability.

It is therefore essential to provide a process for detection of skin color that is adaptative.

There are two possible approaches of coping with this variability problem:
- evaluating the quality of lighting and implementing comprehensive renormalization of images,
- detection of human traits used as reference to identify skin color (example: if a face is recognized, then the pixels on that area can be used as representative of the color of the skin).

This second approach is described in numerous publications. However, for the present application, it is difficult to predict what kind of human element will be apparent in the picture (eye, face, mouth, hands, feet, genitals...).

This means that it is necessary to have previously recognized all these human characteristics in order to determine the color of skin on the image. But in the present method of scoring a pornographic video, the process uses skin color as a first filter tool, directing other software agents or processes to potentially interesting areas. Therefore, a process with previous validation of all the features necessary for recognition of a human, prior to skin color detection, is not suitable for the present application.

The purpose of the process is to achieve automatic evaluation of the quality of lighting and white balance.

To demonstrate the variability of color due to variations of white balance, the inventors used a digital camera to photograph the same scene several times under different settings.

Four video sequences were taken with a commercial digital camcorder, using the four standard default settings for white balance (normal, hold, indoor, outdoor), and this revealed that the color contents of the images are very different.

For instance, a most accurately color corrected indoor image is indeed obtained with a standard "indoor" setting.

It should be further noted that the colors of different zones on an image can vary enormously from one image to another, showing that the white balance dramatically affects and alters the hue. In other words, it is not
possible to recognize colors without addressing the main variability factor of the problem.

This variability has been analyzed in the following manner: in the RGB space, each plot representing a pixel was mapped out in 3 plans (RG, RB, and BG). The y=x line (rectilinear) and the quadratic regression line (curved) were printed, as shown in the graphs of Fig. 4.

These graphs illustrate the fact that "the veil effect" (example: yellow veil) is a non-linear deformation of the cloud of points compared to the diagonal line y = x.

Image renormalization was then tested, in order to correct levels of RGB color so as to reduce the parable trend on the diagonal in order to replicate the straight line trend y = x).

This solution works, renormalizes the color and consequently removes the veil (yellow, blue...). It is then possible to recognize colors with greatly enhanced performance.

However, such renormalization operation significantly lengthens the processing time.

The second solution tested was to "measure" the distortion of the cloud of points, and to provide these measures to a color recognition system.

Pursuant to this testing, this second solution was selected as preferred as it is as efficient and does not require recalculating a re-normalized image.

In order to measure the deformation of the cloud of points, the 3 quadratic regressions (polynomial approximation with a polynomial order 2) were computed and displayed in the graphs of Fig. 4.

This provides 6 coefficients for each image (the coefficient of x² and the coefficient of x, for each of the 3 shots).

These 6 coefficients together constitute what can be considered as a "vector of colorimetric context".

A feed-forward neural network, previously trained to create the color classification (with activation function = hyperbolic tangent and learning rule = back-propagation), is then used.
Before applying the neural network, a rough first selection of eligible areas is computed, eliminating areas too green, too blue, and all black... with a view to optimizing the computation time.

For this purpose, each (R, G, B) pixel is converted into a (H, S, L) pixel and a broad gauge is applied that selects the areas which may be of skin (from orange to pink through red, for a broad range of luminance). It removes full black and for a very large range of saturations - it eliminates the black and white because in this case color is not present).

The membership function for the hue (H) parameter is illustrated in Fig. 5A, and the membership function for the saturation (S) parameter is illustrated in Fig. 5B. A similar membership function is used for the lightness variable (L).

This generates Boolean variables VH, VS and VL for hue, saturation and lightness.

These 3 Boolean variables VH, VS, VL are inputted to a function "LOGIC AND " that decides whether or not the pixel is eligible (0 = "ineligible", 1 = "eligible").

The corresponding block diagram is illustrated in Fig. 6.

The neural network is applied on a thumbnail of p x p pixels, e.g. 5 x 5 pixels, whose center successively assumes all possible locations of eligible pixels in the image, as illustrated in Fig. 7.

As shown on the same figure, the neural network reads the input vector [R, G, B] (3 components each corresponding to a basic color - red, green, blue) for the 5 x 5 pixels, which leads to 5 x 5 x 3 = 75 values, plus the "Colohmetric Context" vector (6 components described above) which gives an indication of the overall hue.

The neural network has been previously trained on a large number (e.g. 1000) of labeled images, i.e. area thumbnails that were identified by the human expert as being human skin.

The distinction between skin color and dark skin color is important to decipher as scenes of nudity that expose genitals commonly show contrasts
between normal skin and dark-skinned elements (example: areoles are darker than the rest of the breast).

After the learning process, the neural network is capable of outputting 3 values each between -1 and 1 (because of the activation function being hyperbolic tangent).

These three output values are post-processed by a set of decision rules.

In order to achieve this every variable is first reduced to between 0 and 1 by linear transformation, i.e. by adding 1 and dividing the result by 2, and is interpreted as the degree of membership to the class, wherein the classes respectively are dark skin color, normal skin color, and no skin color.

These three variables are designated as FPS, FPC, and NPF.

It is quite possible that all these three levels of membership are high, while it makes no logical sense. This generally occurs in cases where a thumbnail color is far removed from those that were encountered while learning the neural network. The neural network then may respond erratically.

A table of fuzzy decision rules is then advantageously used to calculate a score of skin color associated with an index score of reliability. In a preferred embodiment, the index score of reliability is determined by first calculating a value FP as the "FUZZY LOGIC OR" between the values FPS and FPC, and then calculating the absolute gap E between FP and NPF. Finally the "FUZZY LOGIC AND " between E and FP is calculated.

The result of this calculation is identified as the "score of skin color."

This solution has been experimented with satisfactory results in most cases. However, sometimes the neural network learning process has concentrated the neural nets output values mainly between two values -a and +b with $a < 1$ and $b < 1$.

In such cases, it is necessary to reshape the dynamics of these outputs, rather than linearly converting them as explained above.
This dynamics reshaping can be done by applying a fuzzy rule to the outputs of the neural network, as shown in Figs. 8A, 8B and 8C.

It should be noted here that the membership function slope spreads between (average - standard deviation) and (average + standard deviation) of the neurons' output.

The score of skin color is then computed in the same manner as indicated above.

The result of this process is a score of skin color for each pixel of the image, where:

- for ineligible areas, the score is zero,
- for eligible areas, the score is that which is calculated by the process described above.

From these score values, an image is constructed wherein each pixel is replaced by the score at the pixel. This image is called "image of skin color scores."

When the original image is noisy, the image of skin color scores may display irregularities, such as a high score alone surrounded by low scores, or a low score alone surrounded by high scores.

In order to remove the noise of this image and to generate a black and white (binary) image, a mathematics automaton is preferably used and performs the following functions:

- for each pixel, the sum of the pixel and its 8 neighbors as illustrated in Fig. 9 is computed;
- if the sum exceeds a threshold S, then the automaton transforms the pixel to 1 (i.e. gives it the logical value B = 1) and it will, consequently be considered to have a color of skin. Otherwise, the process transforms the pixel to 0 (it is given the logical value B = 0), and it will then considered as not having a color of skin.

At the end of this phase, a Boolean value B is obtained for each pixel.

The entire Boolean image is then analyzed in order to label the regions, a region being a contiguous area where all pixels are equal to B=1.
A set of regions (with their location and size) is obtained. All regions with a size smaller than a given threshold $T_{min}$ are eliminated.

The regions that remain are those considered to have a color of skin.

If the proportion of these regions in the image is greater than a threshold, the image is considered as showing sex-related nudity.

5) Optical Flow process

The present invention advantageously includes an optical flow process as published by Pedro Cobos Arribas and Felix Monasterio Huelin Macia, which offers a real-time version of the Shun & Horn algorithm.

6) Recurrent movement detection

This process is applied after calculating the optical flow. The optical flow can track points that move from one picture to another.

The optical flow provides a vector of apparent movement of every pixel from the previous image to the current image.

This displacement is a 2-component vector computed at each point of the image.

Figs. 10A and 10B show two examples of optical flow, respectively the optical flow of a panoramic view (the pixels move in majority in a given direction) and of a traveling sequence (the pixels move closer to the camera, without common direction).

At each pixel No. $i$, there are two components $D_x$ and $D_y$ which respectively represent the apparent movement among $x$ between image $i$ and image $i + 1$, and apparent movement among $y$ between image $i$ and image $i + 1$.

For a given pixel of the first image of the sequence, the series of $D_x$ when all images in the sequence are successively analyzed is called the signal $S_x$ of apparent travel of that pixel in the sequence. Similarly, the series of $D_y$ when all images in the sequence are successively analyzed is called the signal $S_y$ of apparent travel of that pixel in the sequence.
The process then checks the periodic features of Sx and Sy. The process is the same for Sx and Sy, so that it will be described below with reference to a generic signal S.

If the camera has a linear displacement (example: panoramic, traveling, ..), then S is not stationary.

The first step is to convert S into a stationary signal, i.e. to remove components due to slow or linear movements of the camera.

To do so, the linear regression of S(t) among t is computed, said regression being the equation of the “trend” which is then removed from S(t) to give the stationary signal Sd(t). Sd(t) is called the “trendless signal”.

This is illustrated in Fig. 11.

The next step is to compute the autocorrelation function (FAC) of the Sd signal, as illustrated in Fig. 12:

\[ \text{FAC}(\tau) = \text{correlation} \left( Sd(t), Sd(t-\tau) \right) \]

By definition, the FAC signal passes through its maximal peak at zero.

If S is a periodic signal, then FAC is also a periodic signal. If S is a periodic signal with a noise component, then FAC is a signal that is rapidly becoming periodic when \( \tau \) grows.

The next step is to count the number of crossings of zero of the FAC signal.

The cosine signal CNpz with the same number of points and having the same number of crossings of zero is then generated, and the correlation between FAC(\( \tau \)) and CNpz(\( \tau \)) is computed (cf. Fig. 13).

This correlation C varies in theory between -1 and +1. If it exceeds a threshold K (example: K = 0.75), then it is considered that the signal S is periodic.

The frequency of Sd is the inverse of the period of the cosine CNpz.

When C is higher than the threshold, the frequency F (which allows to determine whether the period may or may not correspond to a type of movement, such as a human movement which cannot exceed certain
speeds) is calculated. Very high frequencies correspond to periodical movements that cannot be generated by a human being).

Likewise, the standard deviation of Sd is considered as the effective amplitude of motion in the analyzed direction.

With the availability of 2 signal components of the movement (among x and y), the frequencies of these two signals are tested independently from each other.

If only one of two components is periodic, then it means that the movement is either horizontal or vertical (depending on which component x or y is periodic).

If the two components are periodic and have the same frequency $F$, then the vector sum of amplitudes gives the direction of the periodic movement in the image.

Movements with amplitude and frequencies within prescribed ranges are considered by the process as illustrating sex-related recurrent movements.

7) Penetration recognition process

"Penetration" is defined as the shape of the torus of contact when a tube is embedded in a hole.

A preliminary analysis of sex scenes has detected that such torus forms are characteristic of pornography (especially "hardcore" pornography). The variability of the torus - as a function of the nature of the penetration - exists, but remains limited.

Typically, the poor quality of the pornographic images (especially in the videos that are highly compressed to be transferred within limited bandwidth) combined with the inherent variability of penetration and the extrinsic variability (conditions of filming: lighting, shadows...) render ineffective the conventional methods of Pattern Recognition, which are based on a segmentation (contours or regions) of the image.
In the process of the present invention, the overall picture is scanned by a Region of Interest (ROI) of a given size (e.g. 32 x 32 pixels) that contains a thumbnail. The ROI then successively assumes all possible locations in the picture, as illustrated in Fig. 1.

Some thumbnails will include torus representative of sexual penetration, and they can be extremely diverse as of actual shape and size.

There are many possible vector bases for the decomposition of an image: complex exponentials, wavelets... These bases all include a wave shape (and therefore represent 'energy') and were designed (example: Fourier Transform) to study linear systems. Indeed, the sinuses (in fact, the complex exponentials) are eigenvectors of linear systems, which makes them particularly relevant signals for the study of these systems.

The uncertainty principle of Heisenberg's theory applied to signal theory dictates that a specific spot localized in a spatial image is diluted by all the components of an energy-based analysis. On the other hand, a general characteristic in the image space is represented at only one point of the frequency space.

But a representation space is needed that does not dilute the details that are of greatest interest to the penetration detection process (those that represent elements to recognize). Also needed is a representation space that is insensitive to noise. To ensure this insensitivity to noise, analyzed signals include redundancy (unlike a base of the vector space signals/images).

The process uses a set of images constructed for this ad-hoc decomposition into basic components. The set of these components is then considered a signature of the analyzed image.

The component of the image among a given thumbnail of decomposition is by definition the coefficient of linear correlation between the analyzed image and the decomposition thumbnail.

A preferred set of decomposition thumbnails is a set of synthetic images that represent each feature of the torus of a possible penetration which the process attempts to recognize.
A set of 14 such thumbnails is illustrated in Fig. 14.

To make the decomposition invariant with respect to rotation, two approaches are preferred:

- from each original decomposition thumbnail, R subsequent thumbnails are deduced through R rotations by 360/R degrees, and are then used for the decomposition of the image on the 14 x R thumbnails. The process then keeps the maximum value of coefficients among subsequent thumbnails as the coefficient for the original thumbnail. This generates 14 coefficients.

- or, the process knows the specific direction of interest in the image (example: because the recurrent movement detection process as described above has detected a periodic movement in this area which goes in a given direction), the process directly computes the coefficient of every original thumbnail rotated in this direction. This also generates 14 coefficients.

These 14 coefficients form a "curve", which can be considered as a "signature" of the ROI image to be analyzed.

(Note: the set of decomposition thumbnails that serve as a basis for decomposition can be extended beyond the 14 images presented above. Furthermore, in certain cases, it may be necessary to add additional thumbnails. If 2 images exhibit the same 14-coefficient signature although they are known as being different, it is clear that the process needs additional decomposition thumbnails.)

The process then discriminates the signature using a feed-forward neural network trained on thousands of labeled images (labeled by a human operator) using a back-propagation learning rule, as illustrated in Fig. 15.

The activation functions of neurons are hyperbolic tangents.

The two outputs are post-processed by a set of decision rules. These outputs vary between -1 and +1. They are processed through a fuzzy set membership function that is used for two purposes:

- to reduce them to between 0 and 1;
- to reshape the dynamic by positioning the slope of the membership function in a range between (average - standard deviation) and (average + standard deviation).

These membership functions are illustrated in Figs. 16A and 16B.

The variables $Ap$ and $D = (Ap - Anp)$ are combined by a «Fuzzy logic AND» (minimum of the values) so as to generate a "Penetration Score":

\[
\text{Penetration Score} = \text{FUZZY LOGIC AND (Ap, D)}
\]

8) Breast nipple recognition process

A preliminary analysis of pornographic scenes has permitted to detect that such forms are characteristic of pornography (both "soft-core" and "hardcore" pornography). The variability of breast nipples exists, but remains limited.

The problematic is generally the same as for tube-hole interface torus detection as described above.

Here again the process uses a previously constructed set of images for ad-hoc decomposition into basic components. The set of these components is then considered a signature of the analyzed image.

The component of the image among a given thumbnail of decomposition is by definition the coefficient of linear correlation between the analyzed image and the decomposition thumbnail.

The set of decomposition thumbnails is a set of synthetic images that represent each feature of breast nipple that the process is attempting to recognize.

An exemplary set of twelve such thumbnails is illustrated in Fig. 17.

Again, to make the decomposition invariant with respect to rotation, the two approaches as described with reference to the penetration detection process can be used.

However, for decomposition original thumbnails that are already invariant by rotation (example: the first, third, fourth and fifth thumbnails
Fig. 17), there is no action to be taken as the coefficient of the decomposition
among such thumbnails is already invariant by rotation.

The 12 coefficients obtained form a "curve", which is the "signature" of
the ROI image to be analyzed.

It should be noted here that the set of decomposition thumbnails that
serve as a basis for decomposition can be extended beyond the 12 images
shown in Fig. 17. In this regard, if two images exhibit the same 12-coefficient
signature although they have been visually identified as being different from
each other, additional decomposition thumbnails can be used.

The signature is then discriminated using a feed-forward neural
network trained on thousands of labeled images (labeled by a human
operator) using back-propagation learning rule, as illustrated in Fig. 18.

The activation functions of neurons are hyperbolic tangents.

The 2 outputs are post-processed by a set of decision rules. These
outputs vary between -1 and +1. They are processed through a fuzzy set
membership function that is used for two things:
- to reduce them to between 0 and 1
- to reshape the dynamic by positioning the slope of the membership
  function in a range between (average - standard deviation) and (average +
  standard deviation).

The membership functions are illustrated in Figs. 19A and 19B.

The variables Abds and D = (Abds - Anbds) are combined by a
«Fuzzy logic AND » (minimum of the values):

Breast nipple Score = FUZZY LOGIC AND (Ap, D)

Advantageously, the method of the invention receives the scores
provided by the different processes, and uses any conventional decision
schemes (weighted sums and thresholds, etc.) to identify whether the image
or the sequence of images is objectionable (and should be filtered out or
accessible through specific procedure) or not.
The present invention is not limited to the described embodiment, but the skilled person will be able to devise many variants and alternate embodiments.

In particular, the general software architecture in terms of ROI (Region Of Interest) iterations and process executions on each ROI can be implemented in different ways, while it is preferred that ROI iteration is mutualized between different processes involving such ROIs.

In addition, the present invention shall also include the different processes as described above and recited in the claims, but considered individually, and any combination of at least two of such processes.
CLAIMS

1. A method for detecting objectionable contents in a bidimensional pixel image, comprising:
   implementing a first process for detecting geometric features of the image,
   implementing a second process for detecting skin colors in the image,
   implementing a third process for detecting object shapes indicative of sexual contents in the image.

2. A method according to claim 1, wherein said bidimensional pixel image belongs to a video sequence, and further comprising:
   implementing a fourth process for detecting recurrent movement in a succession of images.

3. A method according to claim 1 or 2, wherein said first process includes a sub-process for detecting circular patterns, said sub-process comprising the following steps:
   - scanning the image with an region-of-interest thumbnail having a size of X x Y pixels substantially smaller than the starting image, and
   - for each of a plurality of individual thumbnails:
     • calculating a linear correlation coefficient between said thumbnail and the same thumbnail rotated through 90° relative to said thumbnail, and
     • identifying said thumbnail as containing a circular pattern if the calculated linear correlation coefficient exceeds a threshold.

4. A method according to any one of claims 1 to 3, wherein said first process includes a sub-process for detecting areas having axial symmetry in a bidimensional pixel image, said sub-process comprising the following steps:
- scanning the image with an region-of-interest thumbnail having a size of \( X \times Y \) pixels substantially smaller than the starting image, and
- for each of a plurality of individual thumbnails:
  - calculating a linear correlation coefficient between said thumbnail and the same thumbnail rotated through 180° relative to said thumbnail, and
  - identifying said thumbnail as containing a circular pattern if the calculated linear correlation coefficient exceeds a threshold.

5. A method according to claim 3 or 4, wherein \( X = Y \).

6. A method according to claim 3, 4 or 5, wherein \( X \) and \( Y \) have values between 16 and 64.

7. A method according to any one of claims 3 to 6, wherein said first process is performed for all possible thumbnail positions in the image.

8. A method according to any one of claims 1 to 7, wherein each pixel of the image has values in a color space having \( N \) dimensions (such as RGB), and said second process comprises the following steps:
   - scanning the image with an region-of-interest thumbnail having a size of \( X \times Y \) pixels substantially smaller than the starting image, and
   - for each of a plurality of selected individual thumbnails:
     - in each color subspace having a dimension \( M \) lower than \( N \) (RG, RB, GB), measuring the distortion of the cloud of two dimensional points corresponding to the thumbnail pixel values in this color subspace, so as to provide distortion coefficients forming a vector of colorimetric context,
     - applying said vector to a neural network previously trained with labeled images to generate color classification.
9. A method according to claim 8, further comprising, before applying vector to the neural network, the step of eliminating non-eligible areas of the image (too green, too blue, too dark, etc.).

10. A method according to claim 9, wherein said elimination step comprises converting the color pixel values (RGB) into HSL (Hue/Saturation/Lightness) pixel values and applying to said HSL values a gauge eliminating non-skin type hue values, as well as dark pixels and black & white pixels.

11. A method according to claim 8, wherein said neural network outputs degree of membership values among three classes, i.e. normal skin, dark skin, and non-skin.

12. A method according to claim 11, comprising a further step of applying said degree of membership values to a table of fuzzy decision rules for determining a score of skin color associated with an index score of reliability.

13. A method according to claim 12, further comprising a step of generating an image of skin color scores for a plurality of eligible areas of the starting image, and a step of noise removal and binarization to retain as skin areas those of adjacent pixels all have the same binary value.

14. A method according to claim 2 or to any one of claims 3 to 13 taken in its dependence from claim 2, wherein said third process comprises the following steps:
   - for an image of a sequence, providing a vector \((Dx, Dy)\) of apparent movement relative to a previous image of the sequence in different positions of the image,
- for a plurality of image positions, determining apparent travel signal values \((S_x, S_y)\) at that position, made of series of vectors values for successive images of the sequence in that position,

- searching for potential periodic characteristics the apparent travel signal values by:
  - removing signal components corresponding to slow and/or linear movements of the camera,
  - computing an autocorrelation function (FAC) of said signal values after said components removal,
  - generating a trigonometric function \((CNpz)\) of said autocorrelation function,
  - computing a correlation between said autocorrelation function and said trigonometric function,
  - comparing the correlation value with a threshold so as to identify the signal values \((S_x, S_y)\) as periodic, and
  - determining the frequency of the periodicity of the signal values.

15. A method according to claim 14, further comprising the step of determining the amplitudes of the periodic signal values.

16. A method according to claim 14 or 15, wherein the determinations are made independently along two perpendicular directions.

17. A method according to claims 15 and 16 in combination, wherein, if the frequencies are essentially identical along the two directions, then the amplitudes of the signal values along the two directions are combined in a vector sum.
18. A method according to claim 14, wherein the removal step comprises computing a linear regression of S(t) among t, so as to remove the trend component thereof.

19. A method according to any one of claims 1 to 18, wherein said fourth process comprises the following steps:
   - providing a plurality of decomposition thumbnails having a size of X x Y diagrammatically representing different object appearances,
   - scanning the image to form region-of-interest (ROI) thumbnails of image portions having different sizes,
   - computing a linear correlation between said ROI thumbnails and said decomposition thumbnails with different mutual orientations therebetween,
   - generating a signature of each ROI thumbnail made from the linear correlation values obtained with each decomposition thumbnail, and
   - using a neural network previously trained with labeled images containing said object for discriminating between signatures to deliver a membership function discriminating between object containing images and object-free images.

20. A method according to claim 19, wherein said thumbnails are black & white.

21. A method according to claim 19 or 20, comprising post-processing the membership function, said post-processing including dynamics reshaping by repositioning the slope of the membership function.
FIG. 1

Image n°i of the video sequence n°k

ROI assuming successively every possible location

FIG. 2

90 deg rotation

Linear correlation

Score
FIG. 4C

Image 3 R G

$y = 0.0036x^2 + 1.8148x$

Image 3 R B

$y = 0.0019x^2 + 0.5026x$

Image 3 G B

$y = 0.0012x^2 + 0.6903x$
FIG. 4D

Image 4 R G

\[ y = 0.0021x^2 + 0.2411x \]

Image 4 R B

\[ y = 0.0028x^2 + 0.2212x \]

Image 4 G B

\[ y = 0.0038x^2 + 0.0382x \]
FIG. 5A

Membership function of kept H set

1

0

H_\text{min} \quad H_\text{max}

H

FIG. 5B

This generates a Boolean variable VH

Membership function of kept S set

1

0

S_\text{min} \quad S_\text{max}

S
FIG. 9

Sum of 9 scores

FIG. 10A
Optical flow of a panoramic view

FIG. 10B
Optical flow of a traveling sequence
12 - coefficient signature

FIG. 18

Breast Nipple
No

FIG. 19A

Abds

Output Neuron 1

FIG. 19B

Anbds

Output Neuron 2

SUBSTITUTE SHEET (RULE 26)
### INTERNATIONAL SEARCH REPORT

**International application No**

PCT/EP2009/056765

---

**A. CLASSIFICATION OF SUBJECT MATTER**

INV. G06K9/00  G06K9/32

---

**B. RELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

G06K

---

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal

---

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>-/--</td>
<td>14-21</td>
</tr>
</tbody>
</table>

---

**X** Further documents are listed in the continuation of Box C.

**D** See patent family annex.

---

1. **A** document defining the general state of the art which is not considered to be of particular relevance
2. **E** earlier document but published on or after the international filing date
3. **L** document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
4. **O** document referring to an oral disclosure, use, exhibition or other means
5. **P** document published prior to the international filing date but later than the priority date claimed

---

**Date of the actual completion of the international search**

14 August 2009

---

**Date of mailing of the international search report**

12/10/2009

---

**Name and mailing address of the ISA/Authorized officer**

European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel: (+31-70) 340-2040, Fax: (+31-70) 340-3016

Grigorescu, Cosmin
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No</th>
</tr>
</thead>
</table>
**INTERNATIONAL SEARCH REPORT**

**Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. □ Claims Nos.:
   because they relate to subject matter not required to be searched by this Authority, namely:

2. □ Claims Nos.:
   because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. □ Claims Nos.:
   because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

**see additional sheet**

1. □ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. □ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. □ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. □ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

   1, 2, 14-21

**Remark on Protest**

□ The additional search fees were accompanied by the applicant’s protest and, where applicable, the payment of a protest fee.

□ The additional search fees were accompanied by the applicant’s protest but the applicable protest fee was not paid within the time limit specified in the invitation.

□ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet (2)) (April 2005)
This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. **Claims**: 1,2,14-21
   
   Motion-based detection of illicit (adult) content.

2. **Claims**: 3-13
   
   Image-based detection of shapes in illicit (adult) content.