Power Management of Television system by User State Recognition

Noufal Thangal R*, S. Muppudathi Sutha
Sardar Raja College of Engineering, Tirunelveli, Tamil Nadu, India
***Corresponding author: E-Mail: noufalthangalr@gmail.com**

ABSTRACT

Smart Television is one of the smart electronic devices attracting the consumers today. Better user interface is believed to be the prime characteristics of smart devices to as of smart Television. However the ability of a device to analyze its environment is technically the smartness. With the advancement of technologies the power consumption constraints become more important. In today's television, screen is one of the most consumers of power. Moreover, until the system is switched off, the screen remains active even if the user is not watching. To avoid unnecessary power consumption, the screen shall be dimmed or put to sleep if the user is inactive. This paper presents a power management scheme which automatically recognizes whether the user is watching television or not and makes the screen dim if the user is found inactive for long. This is done by recognizing the user's state by analyzing the user's face detected from the image captured by a camera mounted on the television. The user states considered are User-Active and User-Inactive. The resultant screen states are Screen-Active, Screen-Dim and Screen-Sleep.

KEY WORDS: Power Management, Screen, Smart TV.

1. INTRODUCTION

Power management is one of the most important constraints an Engineer considers. This paper presents a method to manage power consumed by a television. The screen of a television is taken under consideration. Generally, the screen of a personal computer is dimmed when the user remains inactive for a while. This is done by measuring how long all the input devices have been idle. This cannot be done in the case of a television because the user is not always expected to use the input device frequently. For instance, if the user is watching a movie which is a few hours long, the input devices may remain idle. Thus the screen may remain active irrespective of the state of the user. This paper presents a new method which manages the power consumed by the screen by dimming it whenever the user is found inactive for long. Here the user state 'Inactive' may refer to either the user is not watching or not present in front of the television.

In low power design of computers, it is the best practice to design processors that avoid dissipation of power in circuitry that is idle. The degree of control over power management by software varies from one processor to another. It is possible for power management to be entirely controlled by hardware; however software may typically initiate it. System activity is the basis of power management when it comes to event-driven applications such as user interfaces. If the time a system remains idle exceeds some threshold, it is likely that the system continues to be idle for an extended time. Based on this phenomenon, a common approach is to initiate a shutdown when the system is idle for a certain length of time.

Several Processors support varying levels of software control over power management. They use either or some of the power saving modes like doze, nap, standby, sleep and hibernate. The doze mode shuts off most of the functional units but keeps the bus snooping enabled. Doze mode maintains data cache coherency. Nap and sleep modes can each be initiated by software, but a hardware handshake is required to protect data cache coherency. The nap mode shuts off bus snooping and can set a timer to automatically wake up the processor. It keeps a phase-locked loop running to permit quick restart of clocking. On the other hand, the sleep mode shuts off the phase –locked loop. Sleep mode stores all the information of the active processes in RAM so that when it wakes up, those processes can proceed. On the other hand, hibernate mode stores those information in ROM and shuts down the whole system. When the system is restarted, all those processes which were running just before hibernation proceed. Among all the power saving modes, hibernate mode consumes the least amount of power.

In a broad sense, the work comes under the field of Human Machine Interface. According to Fitt's law, the interaction system's execution time depends on the easiness of the user to press the button. The channel switching for instance would require the user to lower his head to look first to which button to choose and then press according to the requirement. Next, he should look up the screen to find whether the channel has changed the expected one or not. If not, then he has to look again into the remote controller to give in the right choice of buttons. The frequent looking up and down increases the time consumed for the single purpose of changing the channel.

Recently, some new user interface technologies has arrived such as the space mouse, the Heads Up Display (HUD), the wristwatch type controller, the touch pad, data glove, and the ring type controller. The space mouse works on the motion and position of user's hand taken by its space sensors. The HUD is a display on the screen highlighted in front of the programmes such that the user need not look down on the remote controller to select the button. The data glove can be used to give in the selection by showing some predefined gestures which are detected by the sensors in the glove which identifies the folding of the fingers. Although these controllers improve user experience, it requires the user to hold or wear the device.

Some hand-free technology has also arrived which include the voice control. This uses the speech recognition module to recognize the user's command spoken out. However, there can be certain difficulties which are unavoidable such as the interference of the sound from the TV programmes. Moreover the speech recognition algorithms are not mature enough to recognize the large vocabularies and classify across various dialects. Additionally, the speech recognition module needs to be kept active all the time which increases the power cost in addition to computational cost of the algorithm.

Hand gesture based TV control is another type of hand-free controlling. A camera is used here to capture the scene in front of the TV. From the stream captured, the user is detected and the gesture is recognized and interpreted as control commands. There are two types of gesture based interface which are icon-based interface, and motion-based interface. They are also applied together and adjusted accordingly to support natural hand control. Controlling becomes much easier for the user to master and use, when it comes to hand-free control. Some new sensors capable of depth detection have also been invented which makes the hand gesture recognition more accurate. This makes hand gesture recognition a potential future for TV control. However similar to the speech recognition module used in voice controlled interface, the gesture recognition module also needs to be active all the time in order to detect the gesture which signals the activation command of launching the gesture controlled interface. This results in high computational cost as well as power consumption. Thus, the means to reduce the computational cost or power consumption should be considered carefully.

Generally, with respect to the constraints of the power consumption, two aspects are considered: the sensor device and the recognition module. An RGB (red-green-blue) camera is used generally as sensor in systems of computer vision. A research was done on solution to reduce power consumption. They found a technique by which the gesture recognition system need not run all the time.

Power consumption also occurs in a very high amount through the screen when it is active while the user is inactive. The proposed system is an automatic method of recognizing whether the user is active or inactive based upon which the screen shall be dimmed and later shutdown if the user is still inactive for a predefined time period.
As this system is based on face detection, there should be some valuable face detection algorithms discussed so that the best of them shall be selected based on the requirements and the considered constraints. The constraints considered herewith are the high speed detection and lower power consumption. The following are some of the popular object detection algorithms.

Viola and Jones introduced a machine learning approach for object detection which is capable of processing images extremely rapidly. This algorithm has achieved high detection rates. Their work is nowadays called as Viola Jones algorithm and is distinguished by three key contributions. The first is the introduction of a new image representation called the Integral image which allows the features used by the detector to be computed very quickly. The second is a learning algorithm, based on Ada Boost, which selects a small number of critical visual features from a larger set and yields extremely efficient classifiers. The third contribution is a method for combining increasingly more complex classifiers in a cascade which allows background regions of the image to be quickly discarded while spending more computation on promising object-like regions. The cascade can be viewed as the object specific focus-of-attention mechanism. This approach provides statistical guarantees that the discarded regions are unlikely to have objects of interest. This algorithm does not use image differencing and skin colour detection.

Ojala (2002), introduced a multi resolution approach to gray scale and rotation invariant texture classification based on Local Binary Patterns (LBP) and nonparametric discrimination of sample and prototype distributions. The method is based on the finding that certain local binary patterns termed uniform are fundamental properties of local image texture. Their occurrence histogram is a very powerful texture feature. A generalized gray scale and rotation invariant operator representation is derived which allows for detecting the uniform patterns for any quantization of the angular space and for any spatial resolution. A method for combining multiple operators for multi resolution analysis was also introduced in this work. Their approach was very robust in terms of gray scale variations, since the operator is invariant against any monotonic transformation of the gray scale. Computational simplicity is another advantage. This is because the operator can be realized with a few operations in a small neighborhood and a lookup table.

Dalal and Triggs (2005), introduced a robust visual object recognition algorithm adopting linear Support Vector Machine based human detection as a test case. They proposed that Histograms of Oriented Gradient (HOG) descriptors significantly outperform other feature sets for human detection. They studied the influence of each stage of the computation on performance, concluding that fine-scale gradients, fine orientation binning, relatively coarse spatial binning, and high-quality local contrast normalization in overlapping descriptor blocks are all important for good results. Thus this algorithm is proved to be perfect for detecting human bodies in a scene with a large range of pose variations and backgrounds.

2. EXISTING SYSTEM

There exists a work on automatic user state recognition for low-cost television systems. It is composed of both hardware and embedded software. The hardware component consists of Ultrasonic sensors, RGB camera and depth camera. The embedded software processes the sensor data, and drives the TV control. The researchers of that project introduced a scheme under the name "Automatic User State Recognition (AUSR)". This scheme consists of several steps, for instance, the user state initialization or updating, the device activation or sleeping, the recognition module activation or sleeping, and the user input notification, as shown in the Fig.1. The user's action is detected based on the ultrasonic distance array and camera module, and user state is initialized or updated automatically. The user state indicates whether the user is controlling the TV, the camera device and gesture recognition module will be turned on accordingly, and the user will be informed to begin hand gestures by displaying a message on the screen. Otherwise, the camera device and the gesture recognition module will be turned off or slept in order to save power. When the user is informed to show the gestures, the user shall show the pre-defined gestures in accordance with the requirement. Those gestures are recognized and the control is done accordingly. Based on the user's actions in front of the TV, the user state is classified into four types, i.e. Absent, Other Action, Controlling and Watching.

This existing work however relies more on hardware and thus increases the cost. Further study was done to find that the posture of face alone can be used to recognize the user's state rather than using the depth and ranging along with the RGB camera. A user is expected to be present in a normal range of 10 meters and in an angular range of -45° to +45° from normal line. Thus a software approach is taken to design an algorithm to detect the presence and then the posture of faces in front of the Television.

3. PROPOSED SYSTEM

The proposed system is not directly an advancement of the existing one but a new technology relating to it. The proposed system makes the screen bright or dim based on the user activity, i.e. when the user is watching (active) or not watching (inactive) respectively. This is done by detecting the face of the users and analyzing whether they are looking towards the TV or not. A set of face postures looking towards the camera are used to train a Face Detection object and it is used to detect such faces in front of the TV. Then the user state such as User Active and User Inactive is recognized. Based on the user state, the actions such as Screen Active, Screen Dim and Screen Sleep shall be taken. Thus, Computer Vision technology based on Digital Image Processing is used to find new dimensions of Human Machine interface of a consumer electronic product – the Television.

User state recognition is the most important part of the proposed system. The flow chart of it is given in the figure 2. Here the input is the image frame captured by the camera. The camera is placed on the TV and is pointed towards the users. The input image is analysed to recognize whether the user is Active or Inactive. The next important part is the decision making. According to the user state, the decision shall be taken whether to keep the screen active or to dim. If the user remains inactive for a long time, the screen should be put to sleep.

a) System implementation: As discussed earlier, the aim is to make an algorithm to recognize the state of the user of a Television in order to change the state of the screen. As is the project a computer vision problem solving, the input is obviously the image captured by a camera. Some digital image processing is done in accordance with the requirement. The output is the screen state decided according to what the state of the user is. The proposed system shall be divided into two modules, the first being that of User State Recognition and the second the design of Decision making Algorithm.

Figure.1.Block diagram of the Proposed system

b) User state recognition: The images which come from the camera are processed and the user state is recognized. The defined user states are broadly two – User Active and User Inactive. The basic principle is to detect faces on the scene.

i) Face Detection: As discussed in the literature survey section, there are three renowned algorithms of Face Detection which has been used worldwide. They are:

a) Viola-Jones algorithm – Rapid Object Detection using a Boosted Cascade of Simple Features

b) Ojala-Pietikainen-Maenpaa algorithm – Multi-resolution Gray-scale and Rotation invariant Texture Classification with Local Binary Patterns (LBP).

c) Dalal-Triggs algorithm – Histograms of Oriented Gradients (HOG) for Human Detection.

Viola-Jones algorithm is the most used algorithm for Detecting Face when the speed of detection is a constraint. This work also demands the speed of detection. Therefore, this algorithm is chosen for face detection.

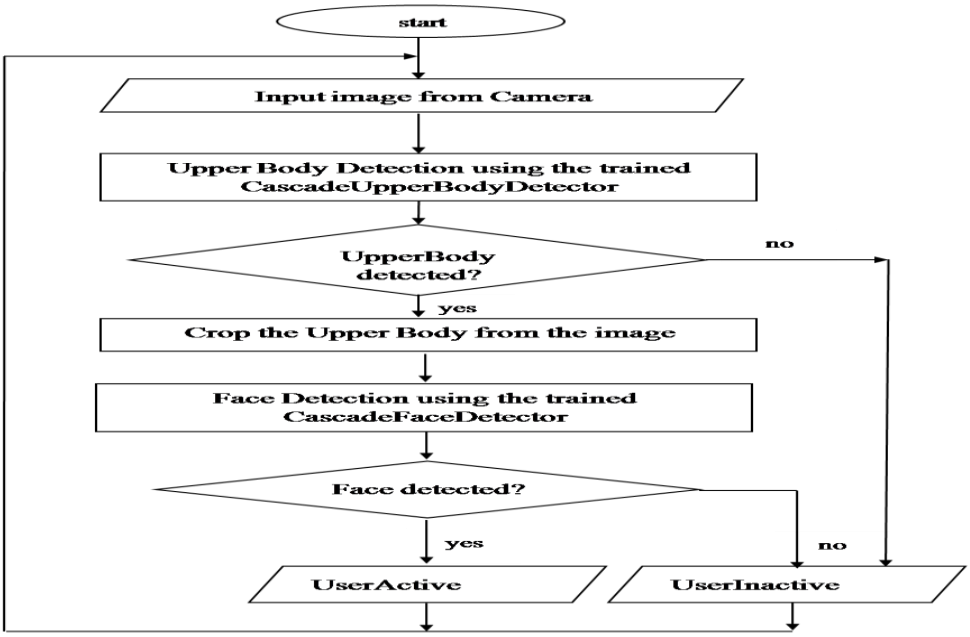

Figure.2.The flow chart of the User State Recognition module

ii) Viola-Jones algorithm: Viola-Jones algorithm is a machine learning approach for visual object detection which is capable of processing images extremely rapidly and achieving high detection rates. Viola-Jones algorithm is characterized by three key contributions such as:

a) A new image representation called Integral Image which allows the features used by the detector to be computed very quickly

b) A learning algorithm, based on Ada Boost, which selects a small number of critical visual features from a larger set and yields extremely efficient classifiers

c) A method for combining increasingly more complex classifiers in a cascade which allows background regions of the image to be quickly discarded while spending more computation on promising object-like regions.

This object detection procedure classifies images based on the value of simple features. The most common reason for using features rather than pixels directly is that features can act to encode ad-hoc domain knowledge that is difficult to learn using finite quantity of training data. Moreover, the feature based system operates much faster than a pixel based system. The simple features used are reminiscent of Haar basis functions. More specifically, three kinds of Features are used. The value of a two-rectangle feature is the difference between the sums of the pixels within two rectangular regions. The regions have the same size and shape and are either horizontally or vertically adjacent. A three-rectangle feature computes the sum within two outside rectangles subtracted from the sum in a centre rectangle. Finally a four-rectangle feature computes the difference between diagonal pairs of rectangles.

Rectangle features can be computed very rapidly using an intermediate representation for the image which we call the integral image. The integral image at location (x,y) contains the sum of the pixels above and to left of (x,y), inclusive:

$$ii(x, y) = \sum_{x' <= x, y' \leq y} i(x', y') \quad (1)$$

Where $ii(x, y)$ is the integral image and $i(x, y)$ is the original image. Using the following pair of recurrences:

$$s(x, y) = s(x, y-1) + i(x, y) \quad (2)$$

$$ii(x, y) = ii(x-1, y) + s(x, y) \quad (3)$$

Where $s(x, y)$ is the cumulative row sum, $s(x, -1) = 0$, and $ii(-1, y) = 0$ the integral image can be computed in one pass over the original image.

The cascade training process involves two types of tradeoffs. In most cases classifiers with more features will achieve higher detection rates and lower false positive rates. A very simple framework is used to produce an effective classifier which is highly efficient. Each stage in the cascade reduces the false positive rate and decreases the detection rate. A target is selected for the minimum reduction in false positives and the maximum decrease in detection. Each stage is trained by adding features until the target detection and false positives rates are met. Stages are added until the overall target for false positive and detection rate is met.

c) Decision Making: The decision making of the algorithm is based on the presence and absence of face in a scene. The decision is to which state the Screen should be kept. This Decision is taken in accordance with the user state taken from the User State Recognition Module. The user states are User Active which means the user is watching the TV and User Inactive which means the user is either absent or is not watching.

The screen should be active as long as the user is active. When the user is inactive, a timer starts counting and if the user remains inactive for a pre-defined time period, the screen shall be dimmed. Then, if the user remains inactive for longer than two periods or a pre-defined period, the screen shall be put to sleep.

At start, the Timer is initialized to zero and the screen is kept active. Then the user state is read in from the User State Recognition module. If the state is User Active, no change occurs, i.e., Timer remains zero and Screen remains active. If the state is User Inactive, Timer is incremented. The timer is then checked for crossing two predefined values such as N and 2N. If the Timer crosses N, then the screen is dimmed and if it crosses 2N, the screen is put to sleep. Else the next user state is taken in. All these repeat for each pulse.

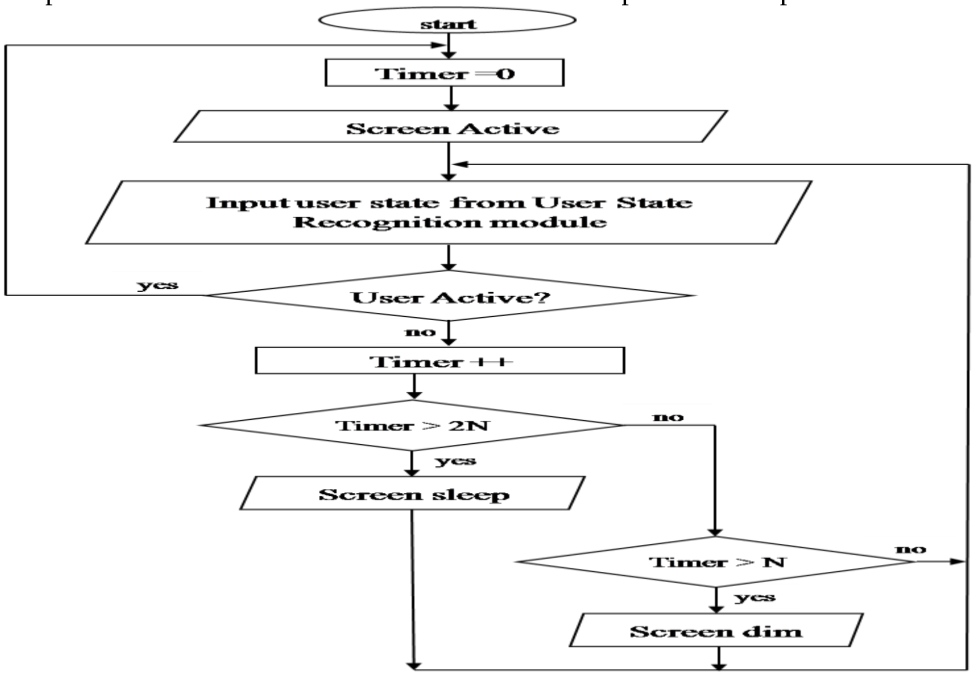

Figure.3.The Flow chart of Decision Making module

Thus there are three decision blocks, one for checking the user state, one for checking Timer cross N, and the other for Timer cross 2N. The only input block is to take in the User State from the User State Recognition module. The output from this module of Decision making is the decisions such as Screen Active, Screen Dim and Screen Sleep. This module thus decides the state of the screen. The screen returns to active state whenever the user appears active.

4. TEST AND OBSERVATIONS

This proposed system relies on the accuracy of the face detection algorithm. There are many cases which are to be considered in a work of this kind. In the face detection algorithm there is a possibility of false positives and false negatives. They can cause issues in the working. If a false positive occur in a scene with no human present, it will result in User Active state which further result in Screen Active which is undesirable. Likewise, a false negative results in User Inactive while there is an active user (the user who watches the TV – face towards the camera).

A test set was taken containing 53 images considering many criteria such as with and without faces, indoor and outdoor locations, active and inactive faces, etc. Each image was subject to User State Recognition process. With the face detected output displayed, the counts of True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) were taken manually. Using these measures, the Accuracy, Sensitivity and Specificity were calculated.

Accuracy measures the proportion of true results in the population.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}} = 85.4651\%$$

Sensitivity measures the proportion of actual positives which are correctly identified as such.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} = 84.8485\%$$

Specificity measures the proportion of actual negatives which are correctly identified as such

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} = 86.3014\%$$

This result is quiet reliable but can be improved in the future works on it.

5. CONCLUSION

The need for a method to manage power consumption of a Television was discussed and concluded that the screen shall be dimmed and further put to sleep if the user is inactive thereby avoiding the power wastage. To meet with this requirement, it was found that the state of the user should be recognized and further power saving action should be taken. The user states considered here are User Active and User Inactive. The power saving actions considered here are Screen Active, Screen Dim and Screen Sleep. Thus a power management system was designed which contains two main parts namely the User State Recognition and the Decision making. The user state recognition module takes images from a camera and analyzes them and finds whether there is any user watching the TV by searching for faces on the image. If there is any face, User Active is taken as the user state, else User Inactive is taken. The Decision making module decides what action to be taken on screen based on the user state recognized by the user state recognition module. If the user state is User Active, the screen state is decided to be Screen Active. For User Inactive, a timer is started and if it reaches a predefined count, then the screen state is decided to be Screen Dim and if it goes longer, the screen state is decided to be Screen Sleep. Whenever the user state becomes User Active, the Screen Active is decided and the timer is reset.

REFERENCES

Aoki R, Ihara M, Maeda A, Kobayashi M, and Kagami S, Expanding kinds of gestures for hierarchical menu selection by unicursal gesture interface, IEEE Trans. Consumer Electronics, 57 (2), 2011, 731-737.

Dalal N, and Triggs B, Histograms of Oriented Gradients for Human Detection, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1, 2005, 886–893.

Freeman W.T, and Weissman C.D, Television control by hand gestures," in Proceeding of IEEE International Workshop on Automatic Face and Gesture Recognition, Zurich, Switzerland, 1995, 179-183.

Han Y.M, A low-cost visual motion data glove as an input device to interpret human hand gestures, IEEE Trans. Consumer Electronics, 2010.

Heo H, Lee E.C, Park K.R, Kim C.J, and Whang M.C, A realistic game system using multi-modal user interfaces, IEEE Trans.Consumer Electronics, 2010.

Jeong S, Jin J, Song T, Kwon K, and Jeon J.W, Single-Camera Dedicated Television Control System using Gesture Drawing, IEEE Trans. on Consumer Electronics, 58 (4), 2012, 1129-1137.

Kim S.K, Park G.H, Yim S.H, Choi S.M, and Choi S.J, Gesture recognizing hand-held interface with vibrotactile feedback for 3D interaction, IEEE Trans. Consumer Electronics, 2009.

Lee D.W, Lim J.M, John S.W, Cho I.Y, and Lee C.H, Actual remote control: a universal remote control using hand motions on a virtual menu, IEEE Trans.Consumer Electronics, 2009.

Liu X, and Fujimura K, Hand gesture recognition using depth data, in Proceedings of 6th IEEE International Conference on Automatic Face and Gesture Recognition, 2004, 529-534.

Mackenzie I.S, and Buxton W, Extending fitts' law to two dimensional tasks, in Proceedings of CHI' 92, 1992, 219-226.

Miranda L.C, Hornung H.H, Baranauskas M.C, Adjustable interactive rings for iDTV, IEEE Trans. on Consumer Electronics, 2010.

Ojala T, Pietikainen M, and Maenpaa T, Multiresolution Gray-scale and Rotation Invariant Texture Classification With Local Binary Patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24 (7), 2002, 971–987.

Papp I, Saric Z, Teslic N, Hands-free voice communication with TV, IEEE Trans. on Consumer Electronics, 2011.

Park J.S, Jang G.J, Kim J.H, Kim S.H, Acoustic interference cancellation for a voice-driven interface in smart TVs, IEEE Trans. On Consumer Electronics, 2013.

Shiguo Lian, Wei Hu, and Kai Wang, Autoamtic User State Recognition for Hand Gesture Based Low-Cost Television Control System, in the IEEE Transactions on Consumer Electronics, 60 (1), 2014.

Takahashi M, Fujii M, Naemura M, and Satoh S, Human gesture recognition using 3.5-dimensional trajectory features for hands-free user interface, in Proceeding of ARTEMIS'10, 2010, 3-8.

Viola P, and Jones M.J, Rapid Object Detection using a Boosted Cascade of Simple Features, Proceedings of the 2001 IEEE Computer Society Conference, 1, 2001, I-511–I-518.