facial image; L_i is the illumination component, representing the illumination and shadow of the facial image, and (x,y) represents the coordinates of the pixels in the image; log-transform the face image I_i, to obtain:

$$\begin{aligned} f_i &= \log I_i \\ &= \log Ri + \log I_i \\ &\triangleq v_i + u_i \end{aligned}$$

where f_i, v_i and u_i respectively represent the value of I, R and L over the log-domain, i.e. v_i=logR, u_i=logL, compute DCT for f_i, i.e.:

$$F_i(s,t) = \frac{1}{4} C_s C_t \sum_{y=0}^{N-1} \sum_{x=0}^{N-1} f_i(x,y) \cos\left(s\pi \frac{2y+1}{2N}\right) \cos\left(t\pi \frac{2x+1}{2N}\right)$$

$$\text{Where } C_s = \begin{cases} 1/\sqrt{2}, & \text{if } s = 0 \\ 1, & \text{Others} \end{cases}, \quad C_t = \begin{cases} 1/\sqrt{2}, & \text{if } t = 0 \\ 1, & \text{Others} \end{cases}$$

where N is the length and width of the image, and the high frequency coefficient of $F_i(s,t)$ is set at 0, i.e.:

$$F_i'(s,t) = \begin{cases} F_i(s,t) & s < M \text{ and } t < M \quad \textit{low frequencies} \\ 0 & s \geq M \text{ and } t \geq M \quad \textit{high frequencies (others)} \end{cases}$$

where M is a parameter to be defined, which is generally set at 5, compute the inverse DCT (discrete cosine transform) for the adjusted frequency domain coefficient F˘, i.e.:

$$f'_i(x,y) = \frac{1}{4} \sum_{s=0}^{N-1} \sum_{t=0}^{N-1} C_s C_t F(s,t) \cos\left(s\pi \frac{2y+1}{2N}\right) \cos\left(t\pi \frac{2x+1}{2N}\right) \quad (6)$$

Take f_i as the estimation of the illumination component, i.e.:

$$u_i(x,y) \approx f_i'(x,y) \quad (7)$$

Then the illumination component of the facial image can be obtained via inverse logarithmic transformation, i.e.:

$$L_i(x,y) = \exp(u_i(x,y))$$

real face.

Preferably, each of the facial images obtained is denoted as I_i, where i is a natural number.

Preferably, the illumination component of each face image I_i, which, according to the Lambertian model, can be denoted as:

$$I_i(x,y)=R_i(x,y)L_i(x,y)$$

where R_i is the reflection component, representing the surface reflectance of the facial image; L_i is the illumination component, representing the illumination and shadow of the facial image, and (x,y) represents the coordinates of the pixels in the image; log-transform the face image I_i, to obtain:

$$\begin{aligned} f_i &= \log I_i \\ &= \log R_i + \log I_i \\ &\triangleq v_i + u_i \end{aligned}$$

where f_i, v_i and u_i respectively represent the value of I, R and L over the log-domain, i.e. $v_i=\log R$, $u_i=\log L$, compute DCT for f_i, i.e.:

$$F_i(s,t) = \frac{1}{4} C_s C_t \sum_{y=0}^{N-1} \sum_{x=0}^{N-1} f_i(x,y) \cos\left(s\pi \frac{2y+1}{2N}\right) \cos\left(t\pi \frac{2x+1}{2N}\right)$$

Where $C_s = \begin{cases} 1/\sqrt{2}, & \text{if } s = 0 \\ 1, & \text{Others} \end{cases}$, $C_t = \begin{cases} 1/\sqrt{2}, & \text{if } t = 0 \\ 1, & \text{others} \end{cases}$

where N is the length and width of the image, and the high frequency coefficient of $F_i(s,t)$ is set at 0, i.e.:

$$F_i'(s,t) = \begin{cases} F_i(s,t) & s < M \text{ and } t < M \quad \textit{low frequencies} \\ 0 & s \geq M \text{ and } t \geq M \quad \textit{high frequencies (others)} \end{cases}$$

where M is a parameter to be defined, which is generally set at 5, compute the inverse DCT (discrete cosine transform) for the adjusted frequency domain coefficient F˘, i.e.:

$$f_i'(x,y) = \frac{1}{4} \sum_{s=0}^{N-1} \sum_{t=0}^{N-1} C_s C_t F(s,t) \cos\left(s\pi \frac{2y+1}{2N}\right) \cos\left(t\pi \frac{2x+1}{2N}\right) \quad (6)$$

$$F_i'(s,t) = \begin{cases} F_i(s,t) & s < M \text{ and } t < M \quad low\ frequencies \\ 0 & s \geq M \text{ and } t \geq M \quad high\ frequencies\ (others) \end{cases}$$

Where M is a parameter to be defined, which is generally set at 5,

Compute the inverse DCT (discrete cosine transform) for the adjusted frequency domain coefficient F˜, i.e.:

$$f'_i(x,y) = \frac{1}{4}\sum_{s=0}^{N-1}\sum_{t=0}^{N-1} C_s C_t F(s,t)\cos\left(s\pi\frac{2y+1}{2N}\right)\cos\left(t\pi\frac{2x+1}{2N}\right)$$

Take f_i as the estimation of the illumination component, i.e.:

$$u_i(x,y) \approx f_i'(x,y)$$

Then the illumination component of the image domain can be obtained via inverse logarithmic transformation, i.e.:

$$L_i(x, y) = \exp(u_i(x, y))$$

Preferably, M has an empirical value of 5. It would be appreciated that M is an adjustable parameter and can take on other values apart from the value 5.

Step 3 requires calculating the mean local variance for the illumination component of the face images obtained from T successive video frames:

Divide the illumination component of each face image equally into aBb image blocks with aBb pixels contained in each block, and use Bij to represent the No.j image block of the face image from the No.i frame, therefore the mean local variance for T successive video frames is:

$$Avar = \frac{1}{T}\sum_{i=1}^{T}\sum_{j=1}^{a\times b} var(B_{i,j})$$

Where $var(B_{i,j})$ is the variance of the pixel value of the image block $B_{i,j}$.

The method for face in-vivo detection based on illumination component, wherein Step 4 requires conducting face in-vivo detection:

$\triangleq v_i + u_i$

Where f_i, v_i and u_i respectively represent the value of I, R and L over the log-domain, i.e. v_i=logR, u_i=logL, By this time, values of v_i and u_i are unknown and the value of u_i needs to be estimated.

Compute DCT (discrete cosine transform) for f_i, i.e.:

$$F_i(s,t) = \frac{1}{4} C_s C_t \sum_{y=0}^{N-1} \sum_{x=0}^{N-1} f_i(x,y) \cos\left(s\pi \frac{2y+1}{2N}\right) \cos\left(t\pi \frac{2x+1}{2N}\right) \quad (3)$$

$$\text{Where } C_s = \begin{cases} 1/\sqrt{2}, & \text{if } s = 0 \\ 1, & \text{Others} \end{cases}, \quad C_t = \begin{cases} 1/\sqrt{2}, & \text{if } t = 0 \\ 1, & \text{Others} \end{cases} \quad (4)$$

Where N is the length and width of the image, and the high frequency coefficient in $F_i(s,t)$ is set at 0, i.e.:

$$F_i'(s,t) = \begin{cases} F_i(s,t) & s < M \text{ and } t < M \quad \textit{low frequencies} \\ 0 & s \geq M \text{ and } t \geq M \quad \textit{high frequencies (others)} \end{cases} \quad (5)$$

Where M is a parameter to be defined, which is generally set at 5,

Compute the inverse DCT (discrete cosine transform) for the adjusted frequency domain coefficient $F^{\sim}$, i.e.:

$$f_i'(x,y) = \frac{1}{4} \sum_{s=0}^{N-1} \sum_{t=0}^{N-1} C_s C_t F(s,t) \cos\left(s\pi \frac{2y+1}{2N}\right) \cos\left(t\pi \frac{2x+1}{2N}\right) \quad (6)$$

Formula from (3) to (6) actually demonstrate the low frequency filtering for the face image fi over the log-domain via DCT (discrete cosine transform).

It would be appreciated that in other embodiments, it can be synchronized with surface shading algorithms, ray cast and the like. It would also be appreciated that M is an adjustable parameter and can take on other values apart from the value 5.

According to a large number of existing researches, as the illumination component in images varies slowly, the low frequency component can be used to estimate the illumination component. Therefore, $f_{\text{Þ}}$ can be used as estimation of the illumination component, i.e.:

of $F_i(s.t)$ is set at 0. i.e.:

$$F_i'(s,t) = \begin{cases} F_i(s,t) & s < M \text{ and } t < M \quad low\ frequencies \\ 0 & s \geq M \text{ and } t \geq M \quad high\ frequencies\ (others) \end{cases}$$

where M is a parameter to be defined, which is generally set at 5, compute the inverse DCT (discrete cosine transform) for the adjusted frequency domain coefficient F˘, i.e.:

$$f_i'(x,y) = \frac{1}{4}\sum_{s=0}^{N-1}\sum_{t=0}^{N-1} C_s C_t F(s,t)\cos\left(s\pi\frac{2y+1}{2N}\right)\cos\left(t\pi\frac{2x+1}{2N}\right) \quad (6)$$

Take f_{i}as the estimation of the illumination component, i.e.:

$$u_i(x,y) \approx f_i'(x,y) \quad (7)$$

Then the illumination component of the facial image can be obtained via inverse logarithmic transformation, i.e.:

$$L_i(x, y) = \exp(u_i(x, y))$$

It would be appreciated that in other embodiments, it can be synchronized with surface shading algorithms, ray cast and the like. It would also be appreciated that M is an adjustable parameter and can take on other values apart from the value 5.

The step of calculating the mean local variance for the illumination components of the face images obtained from T successive video frames comprises:

dividing the illumination component of each face image equally into aBb image blocks with aBb pixels contained in each block, and use Bij to represent the No.j image block of the face image from the No.i frame, therefore the mean local variance for T successive video frames is:

$$Avar = \frac{1}{T}\sum_{i=1}^{T}\sum_{j=1}^{a\times b} var(B_{i,j})$$

where $var(B_{i,j})$ is the variance of the pixel valufe of the image block $B_{i,j}$.

In the step of comparing the mean local variance (Avar) obtained with a predetermined threshold (Th), if the Avar value is greater than or equal to Th, the face image in the video is an image of a real face. If the Avar value is less than Th, the face image in the

CLAIMS

1. A method for face detection, comprising:

capturing a facial movement by video and processing the video to obtain a plurality of facial images from a plurality of successive video frames;

using the Lambertian model to render each facial image obtained and computing discrete cosine transform (DTC) to obtain an illumination component of each facial image;

calculating the mean local variance for the illumination components of the facial images; and

comparing the mean local variance with a predetermined threshold (Th) to determine whether the facial image is an image of a real face.

2. The method according to claim 1, wherein each of the facial images obtained is denoted as I_i, where i is a natural number.

3. The method according to claim 2, wherein the illumination component of each face image I_i, which, according to the Lambertian model, can be denoted as:

$$I_i(x,y)=R_i(x,y)L_i(x,y)$$

where R_i is the reflection component, representing the surface reflectance of the facial image; L_i is the illumination component, representing the illumination and shadow of the facial image, and (x,y) represents the coordinates of the pixels in the image; log-transform the face image I_i, to obtain:

$$f_i = \log I_i = \log R_i + \log I_i \triangleq v_i + u_i$$

where f_i, v_i and u_i respectively represent the value of I, R and L over the log-domain, i.e. $v_i=\log R$, $u_i=\log L$, compute DCT for f_i, i.e.:

$$F_i(s,t) = \frac{1}{4} C_s C_t \sum_{y=0}^{N-1} \sum_{x=0}^{N-1} f_i(x,y) \cos\left(s\pi \frac{2y+1}{2N}\right) \cos\left(t\pi \frac{2x+1}{2N}\right)$$

$$\text{Where } C_s = \begin{cases} 1/\sqrt{2}, & \text{if } s = 0 \\ 1, & \text{Others} \end{cases}, \quad C_t = \begin{cases} 1/\sqrt{2}, & \text{if } t = 0 \\ 1, & \text{Others} \end{cases}$$

where N is the length and width of the image, and the high frequency coefficient of $F_i(s,t)$ is set at 0, i.e.:

$$F_i'(s,t) = \begin{cases} F_i(s,t) & s < M \text{ and } t < M \quad \textit{low frequencies} \\ 0 & s \geq M \text{ and } t \geq M \quad \textit{high frequencies (others)} \end{cases}$$

where M is a parameter to be defined, which is generally set at 5, compute the inverse DCT (discrete cosine transform) for the adjusted frequency domain coefficient $F^{\sim}$, i.e.:

$$f_i'(x,y) = \frac{1}{4} \sum_{s=0}^{N-1} \sum_{t=0}^{N-1} C_s C_t F(s,t) \cos\left(s\pi \frac{2y+1}{2N}\right) \cos\left(t\pi \frac{2x+1}{2N}\right) \quad (6)$$

Take f_i' as the estimation of the illumination component, i.e.:

$$u_i(x,y) \approx f_i'(x,y) \quad (7)$$

Then the illumination component of the facial image can be obtained via inverse logarithmic transformation, i.e.:

$$L_i(x,y) = \exp(u_i(x,y))$$

4. The method according to claim 3, wherein M has an empirical value of 5.

5. The method according to claim 4, wherein the step of calculating the mean local variance for the illumination components of the face images obtained from T successive video frames comprises:

dividing the illumination component of each face image equally into aBb image blocks with aBb pixels contained in each block, and use Bij to represent the No.j image block of the face image from the No.i frame, therefore the mean local variance for T successive video frames is:

$$Avar = \frac{1}{T}\sum_{i=1}^{T}\sum_{j=1}^{a\times b} var(B_{i,j})$$

where $var(B_{i,j})$ is the variance of the pixel valufe of the image block $B_{i,j}$.

6. The method according to claim 5, wherein in the step of comparing the mean local variance (Avar) obtained with a predetermined threshold (Th), if the Avar value is greater than or equal to Th, the face image in the video is an image of a real face.

7. The method according to claim 6, wherein if the Avar value is less than Th, the face image in the video is not an image of a real face.

8. The method according to any one of the preceding claims, wherein the threshold (Th) is set according to specific image quality, whereby a lower image resolution means a lower threshold Th.

9. A system for face detection, comprising:

an acquisition unit configured to capture a facial movement by video and to process the video to obtain a plurality of facial images from a plurality of successive video frames;

a calculation unit configured to render each facial image obtained using the Lambertian model, compute discrete cosine transform (DTC) to obtain an illumination component of each facial image, and calculate the mean local variance for the illumination components of the facial images; and

a determination unit configured to compare the mean local variance with a predetermined threshold (Th) to determine whether the facial image is an image of a real face.

10. The system according to claim 9, wherein each of the facial images obtained is

denoted as I_i, where i is a natural number.

11. The system according to claim 10, wherein the illumination component of each face image I_i, which, according to the Lambertian model, can be denoted as:

$$I_i(x,y)=R_i(x,y)L_i(x,y)$$

where R_i is the reflection component, representing the surface reflectance of the facial image; L_i is the illumination component, representing the illumination and shadow of the facial image, and (x,y) represents the coordinates of the pixels in the image; log-transform the face image I_i, to obtain:

$$\begin{aligned} f_i &= \log I_i \\ &= \log R_i + \log I_i \\ &\triangleq v_i + u_i \end{aligned}$$

where f_i, v_i and u_i respectively represent the value of I, R and L over the log-domain, i.e. $v_i=\log R$, $u_i=\log L$, compute DCT for f_i, i.e.:

$$F_i(s,t) = \frac{1}{4} C_s C_t \sum_{y=0}^{N-1} \sum_{x=0}^{N-1} f_i(x,y) \cos\left(s\pi \frac{2y+1}{2N}\right) \cos\left(t\pi \frac{2x+1}{2N}\right)$$

Where $C_s = \begin{cases} 1/\sqrt{2}, & \text{if } s = 0 \\ 1, & \text{Others} \end{cases}$, $C_t = \begin{cases} 1/\sqrt{2}, & \text{if } t = 0 \\ 1, & \text{others} \end{cases}$

where N is the length and width of the image, and the high frequency coefficient of $F_i(s,t)$ is set at 0, i.e.:

$$F_i'(s,t) = \begin{cases} F_i(s,t) & s < M \text{ and } t < M \quad low\ frequencies \\ 0 & s \geq M \text{ and } t \geq M \quad high\ frequencies\ (others) \end{cases}$$

where M is a parameter to be defined, which is generally set at 5, compute the inverse DCT (discrete cosine transform) for the adjusted frequency domain coefficient F˜, i.e.:

$$f_i'(x,y) = \frac{1}{4} \sum_{s=0}^{N-1} \sum_{t=0}^{N-1} C_s C_t F(s,t) \cos\left(s\pi \frac{2y+1}{2N}\right) \cos\left(t\pi \frac{2x+1}{2N}\right) (6)$$

Take f_i' as the estimation of the illumination component, i.e.:

$$u_i(x,y) \approx f_i'(x,y) (7)$$

Then the illumination component of the facial image can be obtained via inverse logarithmic transformation, i.e.:

$$L_i(x, y) = \exp(u_i(x, y))$$

12. The system according to claim 11, wherein M has an empirical value of 5.

13. The system according to claim 12, wherein the step of calculating the mean local variance for the illumination components of the face images obtained from T successive video frames comprises:

dividing the illumination component of each face image equally into aBb image blocks with aBb pixels contained in each block, and use Bij to represent the No.j image block of the face image from the No.i frame, therefore the mean local variance for T successive video frames is:

$$\mathrm{Avar} = \frac{1}{T}\sum_{i=1}^{T}\sum_{j=1}^{a\times b}\mathrm{var}(B_{i,j})$$

where $\mathrm{var}(B_{i,j})$ is the variance of the pixel valufe of the image block $B_{i,j}$.

14. The system according to claim 13, wherein in the step of comparing the mean local variance (Avar) obtained with a predetermined threshold (Th), if the Avar value is greater than or equal to Th, the face image in the video is an image of a real face.

15. The system according to claim 14, wherein if the Avar value is less than Th, the face image in the video is not an image of a real face.

16. The system according to any one of claims 9 to 15, wherein the threshold (Th) is set according to specific image quality, whereby a lower image resolution means a lower threshold Th.