

and the multi-frames with different colors are of the entire human face, as opposed to just the eye pupil area only (i.e. eye pupil area images A and B).

[0034] For the multi-frames with different colors, features are extracted using a CNN network (N2). The CNN network N2 can be trained using large scale multi-frames with different colors on both live and spoof faces. In an implementation, resnet18 is adopted as the network structure to train the binary classifier N2.

[0035] A confidence score S2 is computed based on the features that are extracted from the multi-frames with different colors.

[0036] The similarity score S1 and the confidence score S2 can be fused to obtain a final authentication result. A decision function can be defined as: If $S1 < T1$ or $S2 > T2$, the result is a live face, otherwise, the result is a spoof face. The pre-determined thresholds T1 and T2 can be calculated based on a validation dataset.

[0037] In summary, the authentication method can be split into two phases. The first “flash light” phase involves obtaining a similarity score S1 based on features that are extracted from eye pupil area image A (without flash light projected onto user’s eye pupil area) and B (with flash light projected onto user’s eye pupil area). The second “color sequence” phase involves obtaining a confidence score S2 based on features that are extracted from the multi-frames with different colors projected onto the user’s face.

[0038] The color sequence phase can effectively detect most spoof attacks, such as high resolution photograph/screenshot and 2D paper mask. However, for 3D masks, the color sequence phase may not work well since the surface or material of the 3D mask can be quite similar to a human face. During the “flash light” phase when a flash light (very bright white light) is projected onto the human eye, especially the eye pupil area, a clear difference can be observed from the two frames (i.e. with/without flashlight), but the difference may not be found if a 3D mask is used for spoofing. This is because the material of the eye pupil area of the human and 3D mask is different. For more robust fraud detection and to cover more types of attacks, similarity score S1 and confidence score S2 are fused together to make a final decision.

[0055] The method further includes using the trained supplementary CNN to extract a plurality of supplementary features associated with the plurality of supplementary face images of the user. Thereafter, the confidence score is generated based on the plurality of extracted supplementary features.

[0056] The similarity score being less than a pre-determined similarity threshold represents that there is the difference between the extracted first features and the extracted second features. The similarity threshold and the confidence threshold can be separately determined based on a validation dataset. The validation dataset comprises both live face images and spoof face images, and a receiver operating characteristic (ROC) curve can be calculated from the label of validation data and prediction results. In an implementation, from the ROC curve, the threshold is set when the FAR (false accepted rate) is equal to 0.01 or 0.001.

[0057] Figure 3 shows a schematic diagram of a computer system suitable for use in executing at least some steps of the user authentication method.

[0058] The following description of the computer system / computing device 300 is provided by way of example only and is not intended to be limiting.

[0059] As shown in Figure 3, the example computing device 300 includes a processor 304 for executing software routines. Although a single processor is shown for the sake of clarity, the computing device 300 may also include a multi-processor system. The processor 304 is connected to a communication infrastructure 306 for communication with other components of the computing device 300. The communication infrastructure 306 may include, for example, a communications bus, cross-bar, or network.

[0060] The computing device 300 further includes a main memory 308, such as a random access memory (RAM), and a secondary memory 310. The secondary memory 310 may include, for example, a hard disk drive 312 and/or a removable storage drive 314, which may include a magnetic tape drive, an optical disk drive, or the like. The removable storage drive 314 reads from and/or writes to a removable storage unit 318 in a well-known manner. The removable storage unit 318 may include a magnetic tape, optical disk, or the like, which is read by and written to by removable storage drive 314. As will be appreciated by persons skilled in the relevant art(s), the removable storage unit 318 includes a computer readable storage medium having stored therein computer executable program code instructions and/or data.

[0073] A plurality of supplementary face images of the user can be captured whilst a plurality of non-white lights are sequentially projected onto the user's face. The extracting device 402 can extract, using the trained supplementary CNN, a plurality of supplementary features associated with the plurality of supplementary face images of the user, and the score generating device can generate the confidence score based on the plurality of extracted supplementary features.

[0074] The similarity score being less than a pre-determined similarity threshold represents that there is the difference between the extracted first features and the extracted second features. The similarity threshold and the confidence threshold are separately determined based on a validation dataset.

[0075] This specification uses the term "configured to" in connection with systems, devices, and computer program components. For a system of one or more computers to be configured to perform particular operations or actions means that the system has installed on it software, firmware, hardware, or a combination of them that in operation cause the system to perform the operations or actions. For one or more computer programs to be configured to perform particular operations or actions means that the one or more programs include instructions that, when executed by data processing apparatus, cause the apparatus to perform the operations or actions. For special-purpose logic circuitry to be configured to perform particular operations or actions means that the circuitry has electronic logic that performs the operations or actions.

[0076] It will be appreciated by a person skilled in the art that numerous variations and/or modifications may be made to the present invention as shown in the specific embodiments without departing from the spirit or scope of the invention as broadly described. The present embodiments are, therefore, to be considered in all respects to be illustrative and not restrictive.

CLAIMS

1. A user authentication method, comprising:

extracting, using a trained convolutional neural network (CNN), a first feature associated with a first eye pupil image of a user, wherein the first eye pupil image of the user is captured in the absence of a white light being projected onto the user's face;

extracting, using the trained CNN, a second feature associated with a second eye pupil image of the user, wherein the second eye pupil image of the user is captured whilst the white light is projected onto the user's face;

generating a similarity score based on the extracted first feature and the extracted second feature; and

positively authenticating the user on a condition that the similarity score indicates a difference between the extracted first feature and the extracted second feature, wherein the difference is associated with a presence of a white dot in and/or around an eye pupil area whilst the flash light is projected onto the user's face.

2. The method as claimed in claim 1, further comprising:

extracting, using a trained supplementary convolutional neural network (CNN), a supplementary feature associated with a supplementary face image of the user, wherein the supplementary face image of the user is captured whilst a non-white light is projected onto the user's face;

generating a confidence score based on the extracted supplementary feature; and

positively authenticating the user on a condition that (i) the similarity score indicates the difference between the extracted first feature and the extracted second feature or (ii) the confidence score is greater than a pre-determined confidence threshold.

3. The method as claimed in claim 1 or 2, further comprising:

capturing, using an image sensor, a first face image of the user comprising the first eye pupil image;

applying an eye detection method to the first face image of the user to generate a first eye bounding box wherein an area of the first face image of the user within the first eye bounding box corresponds to the first eye pupil image;

capturing, using the image sensor, a second face image of the user comprising the second eye pupil image; and

applying the eye detection method to the second face image of the user to generate a second eye bounding box wherein an area of the second face image of the user within the second eye bounding box corresponds to the second eye pupil image.

4. The method as claimed in any one of the preceding claims, wherein the CNN is trained using a large dataset of eye pupil images comprising eye pupils having the white light projected thereon, and wherein the CNN adopts a resnet18 network structure.

5. The method as claimed in claim 2, wherein the supplementary CNN is trained using a large dataset of face images comprising live face images and spoof face images having the non-white light projected thereon, and wherein the supplementary CNN adopts a resnet18 network structure.

6. The method as claimed in claim 2, wherein a plurality of supplementary face images of the user are captured whilst a plurality of non-white lights are sequentially projected onto the user's face, and wherein the method further comprises:

extracting, using the trained supplementary CNN, a plurality of supplementary features associated with the plurality of supplementary face images of the user; and

generating the confidence score based on the plurality of extracted supplementary features.

7. The method as claimed in any one of the preceding claims, wherein the similarity score being less than a pre-determined similarity threshold represents that there is the difference between the extracted first feature and the extracted second feature, and wherein the similarity threshold and the confidence threshold are separately determined based on a validation dataset.

8. A user authentication system, comprising:
an extracting device configured to:

extract, using a trained convolutional neural network (CNN), a first feature associated with a first eye pupil image of a user, wherein the first eye pupil image of the user is captured in the absence of a white light being projected onto the user's face; and

extract, using the trained CNN, a second feature associated with a second eye pupil image of the user, wherein the second eye pupil image of the user is captured whilst the white light is projected onto the user's face;

a score generating device configured to generate a similarity score based on the extracted first feature and the extracted second feature; and

an authenticating device configured to positively authenticate the user on a condition that the similarity score indicates a difference between the extracted first feature and the extracted second feature, wherein the difference is associated with a presence of a white dot in and/or around an eye pupil area whilst the flash light is projected onto the user's face.

9. The system as claimed in claim 8, wherein
the extracting device is further configured to:

extract, using a trained supplementary convolutional neural network (CNN), a supplementary feature associated with a supplementary face image of the user, wherein the supplementary face image of the user is captured whilst a non-white light is projected onto the user's face;

the score generating device is further configured to generate a confidence score based on the extracted supplementary feature; and

the authenticating device is further configured to positively authenticate the user on a condition that (i) the similarity score indicates the difference between the extracted first feature and the extracted second feature or (ii) the confidence score is greater than a pre-determined confidence threshold.

10. The system as claimed in claim 8 or 9, further comprising an image sensor configured to:

capture a first face image of the user comprising the first eye pupil image;

apply an eye detection method to the first face image of the user to generate a first eye bounding box wherein an area of the first face image of the user within the first eye bounding box corresponds to the first eye pupil image;

capture a second face image of the user comprising the second eye pupil image; and

apply the eye detection method to the second face image of the user to generate a second eye bounding box wherein an area of the second face image of the user within the second eye bounding box corresponds to the second eye pupil image.

11. The system as claimed in any one of claims 8 to 10, wherein the CNN is trained using a large dataset of eye pupil images comprising eye pupils having the white light projected thereon, and wherein the CNN adopts a resnet18 network structure.

12. The system as claimed in claim 9, wherein the supplementary CNN is trained using a large dataset of face images comprising live face images and spoof face images having the non-white light projected thereon, and wherein the supplementary CNN adopts a resnet18 network structure.

13. The system as claimed in claim 9, wherein a plurality of supplementary face images of the user are captured whilst a plurality of non-white lights are sequentially projected onto the user's face, and wherein

the extracting device is further configured to extract, using the trained supplementary CNN, a plurality of supplementary features associated with the plurality of supplementary face images of the user, and

the score generating device is further configured generate the confidence score based on the plurality of extracted supplementary features.

14. The system as claimed in any one claims 8 to 13, wherein the similarity score being less than a pre-determined similarity threshold represents that there is the difference between the extracted first feature and the extracted second feature, and wherein the similarity threshold and the confidence threshold are separately determined based on a validation dataset.