

A USER AUTHENTICATION METHOD AND SYSTEM

TECHNICAL FIELD

[001] The present specification relates broadly, but not exclusively, to a user authentication method and to a user authentication system.

BACKGROUND

[002] electronic-Know Your Customer (eKYC) is a digital due diligence process performed by a business entity or service provider to verify the identity of its customers to prevent falsification of identity. eKYC processes typically include a series of checks conducted at the initial stage of the relationship with the customers to verify their identities.

[003] In a typical eKYC process, a customer is required to take a photograph of his/her face. Face anti-spoofing methods are implemented to prevent false facial verification when attackers use a photograph, high resolution screenshot, 2D/3D mask or other substitute for a person's live face.

[004] Current face anti-spoofing methods may be effective in detecting photographs, screenshots or 2D mask spoof faces. However, current face anti-spoofing methods may not be effective in detecting 3D mask spoof faces.

[005] A need therefore exists to improve the manner in which spoof faces can be detected.

SUMMARY

[006] Embodiments seek to provide a user authentication method which involves a flash light based face anti-spoofing method to detect photographs, screenshots, 2D and 3D mask spoof faces. In the case of a live face, when a flash light is projected onto an eye pupil area, a white dot is expected to be seen in and/or around the eye pupil area. However, in the case of a photograph, screenshot, 2D or 3D mask spoof face, a white dot is not expected to be seen in and/or around the eye pupil area.

[007] According to one embodiment, there is provided a user authentication method, including: extracting, using a trained convolutional neural network (CNN), a first feature associated with a first eye pupil image of a user; 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 a 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 authenticating the user on a condition that the similarity score indicates a difference between the extracted first feature and the extracted second feature.

[008] According to another embodiment, there is provided a user authentication system, including: 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; 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 a 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 authenticate the user on a condition that the similarity score indicates a difference between the extracted first feature and the extracted second feature.

BRIEF DESCRIPTION OF THE DRAWINGS

[009] Embodiments are provided by way of example only, and will be better understood and readily apparent to one of ordinary skill in the art from the following written description, read in conjunction with the drawings, in which:

[0010] Figure 1 is a flow chart illustrating an example of a user authentication method, according to an embodiment.

[0011] Figure 2a shows an example of an eye pupil area image without a flash light projected onto the eye; Figures 2b, 2c and 2d show examples of an eye pupil area image (of a live face) with a flash light projected onto the eye.

[0012] Figure 3 shows a schematic diagram of a computer system suitable for use in executing at least some steps of the user authentication method.

[0013] Figure 4 is a schematic diagram illustrating an example of a user authentication system, according to an embodiment.

DETAILED DESCRIPTION

[0014] Embodiments will be described, by way of example only, with reference to the drawings. Like reference numerals and characters in the drawings refer to like elements or equivalents.

[0015] Some portions of the description which follows are explicitly or implicitly presented in terms of algorithms and functional or symbolic representations of operations on data within a computer memory. These algorithmic descriptions and functional or symbolic representations are the means used by those skilled in the data processing arts to convey most effectively the substance of their work to others skilled in the art. An algorithm is here, and generally, conceived to be a self-consistent sequence of steps leading to a desired result. The steps are those requiring physical manipulations of physical quantities, such as electrical, magnetic or optical signals capable of being stored, transferred, combined, compared, and otherwise manipulated.

[0016] Unless specifically stated otherwise, and as apparent from the following, it will be appreciated that throughout the present specification, discussions utilizing terms such as “receiving”, “scanning”, “calculating”, “determining”, “replacing”, “generating”, “initializing”, “outputting”, or the like, refer to the action and processes of a computer system, or similar electronic device, that manipulates and transforms data represented as physical quantities within the computer system into other data similarly represented as physical quantities within the computer system or other information storage, transmission or display devices.

[0017] The present specification also discloses apparatus for performing the operations of the methods. Such apparatus may be specially constructed for the required purposes, or may comprise a computer or other device selectively activated or reconfigured by a computer program stored in the computer. The algorithms and displays presented herein are not inherently related to any particular computer or other apparatus. Various machines may be used with programs in accordance with the teachings herein. Alternatively, the construction of more specialized apparatus to perform the required method steps may be appropriate. The structure of a computer

suitable for executing the various methods / processes described herein will appear from the description below.

[0018] In addition, the present specification also implicitly discloses a computer program, in that it would be apparent to the person skilled in the art that the individual steps of the method described herein may be put into effect by computer code. The computer program is not intended to be limited to any particular programming language and implementation thereof. It will be appreciated that a variety of programming languages and coding thereof may be used to implement the teachings of the disclosure contained herein. Moreover, the computer program is not intended to be limited to any particular control flow. There are many other variants of the computer program, which can use different control flows without departing from the spirit or scope of the invention.

[0019] Furthermore, one or more of the steps of the computer program may be performed in parallel rather than sequentially. Such a computer program may be stored on any computer readable medium. The computer readable medium may include storage devices such as magnetic or optical disks, memory chips, or other storage devices suitable for interfacing with a computer. The computer readable medium may also include a hard-wired medium such as exemplified in the Internet system, or wireless medium such as exemplified in the GSM mobile telephone system. The computer program when loaded and executed on such a computer effectively results in an apparatus that implements the steps of the preferred method.

[0020] electronic-Know Your Customer (eKYC) is a digital due diligence process performed by a business entity or service provider to verify the identity of its customers to prevent falsification of identity. Authentication can be considered as a form of fraud detection in which legitimacy of users are verified and potential fraudsters may be detected before fraudulent acts are carried out. Effective authentication can enhance data security of systems so that digital data is protected from unauthorized users.

[0021] The techniques described in the present specification produce one or more technical effects. In particular, by computing a similarity/difference between a first eye pupil image of a user and a second eye pupil image of the user (wherein the second eye pupil image of the user is captured whilst a white light is projected onto the user's face), the user authentication method and system may reduce the success rate of attacks to the eKYC process, and may be especially effective identifying attacks that

use a 3D mask. An attack to the eKYC process can be identified if the first eye pupil image of the user and the second eye pupil image of the user are determined to be similar or identical.

[0022] Further, the user authentication method and system may provide even higher accuracy in detecting attacks by computing a confidence score based on extracted supplementary features associated with face image(s) of the user (wherein one or more face image(s) of the user are captured whilst one or more non-white lights are projected onto the user's face).

[0023] Embodiments seek to provide a user authentication method which involves a flash light based face anti-spoofing method to detect photographs, screenshots, 2D and 3D mask spoof faces. In the case of a live face, when a flash light is projected onto an eye pupil area, a white dot is expected to be seen in and/or around the eye pupil area. However, in the case of a photograph, screenshot, 2D or 3D mask spoof face, a white dot is not expected to be seen in and/or around the eye pupil area.

[0024] An image capturing device (e.g. RGB camera) is used to capture a user's face image, without any flash light projected onto the user's face. The RGB camera is equipped with a CMOS sensor through which images of the user's face (or whatever an attacker is presenting for authentication, e.g. photograph, high resolution screenshot, 2D/3D printed mask, etc.) are acquired. The eye area of the captured user's face image is cropped out such that there is greater focus on the eye pupil area. The cropped out eye pupil area image (labelled as "A") is collected. In image "A", no flash light is projected onto the user's face/eye.

[0025] The eye area can be cropped out by first extracting facial points, including eye points. "*Facial Landmark Detection: a Literature Survey*" by Yue Wu, et al, 15 May 2018, discloses techniques of extracting facial points. Thereafter, the eye area / eye pupil area is cropped out based on the eye points. For example, by obtaining the leftmost eye point, rightmost eye point, the uppermost eye point, and the lowermost eye point, it is possible to crop a rectangle area which contains one eye. The eye pupil area can be cropped out in a similar way.

[0026] The image capturing device is used to capture the user's face image whilst a flash light is projected onto the user's face. The eye area of the captured user's face

image is cropped out such that there is greater focus on the eye pupil area. The corresponding eye pupil area image (labelled as “B”) is collected.

[0027] It is expected that a white dot can be clearly seen in the eye pupil area of image B (with flash light). On the other hand, a white dot is not expected to be seen in the eye pupil area of image A (without flash light).

[0028] For each eye pupil area image A and B, features are extracted using a CNN network (N1). The CNN network N1 can be trained using large scale pupil area images with and without a flash light projected onto the eye. In an implementation, resnet18 is adopted as the network structure to train the feature extractor.

[0029] A similarity score S1 is computed based on the features that are extracted from eye pupil area images A and B. In an implementation, to compute the similarity of two feature vectors, a cosine similarity method is adopted. Cosine similarity measures the similarity between two vectors of an inner product space. It is measured by the cosine of the angle between two vectors and determines whether two vectors are pointing in roughly the same direction. A formula that can be used to compute cosine similarity is as follows.

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}},$$

[0030]

[0031] where A_i and B_i are components of vector A and B, respectively.

[0032] If the similarity score S1 is larger than a pre-determined threshold (T1), it implies that the two images A and B are similar. In other words, the result indicates that image B is a spoof image (e.g. the image B is of a 3D mask or screenshot), since the white dot in the eye pupil area is not observed. On the other hand, if the score S1 is smaller than the pre-determined threshold (T1), it implies the two images A and B are different. In other words, it is expected that the white dot in the eye pupil area is observed in image B.

[0033] In an implementation, for more robust fraud detection and greater accuracy, additional steps are performed which involve projecting a sequence of colored lights to the user’s face to obtain multi-frames with different colors (such as, but not limited to, red, blue, yellow, green). The colored lights are projected onto the whole human face,

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 S_2 is computed based on the features that are extracted from the multi-frames with different colors.

[0036] The similarity score S_1 and the confidence score S_2 can be fused to obtain a final authentication result. A decision function can be defined as: If $S_1 < T_1$ or $S_2 > T_2$, the result is a live face, otherwise, the result is a spoof face. The pre-determined thresholds T_1 and T_2 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 S_1 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 S_2 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 S_1 and confidence score S_2 are fused together to make a final decision.

[0039] Figure 1 is a flow chart 100 illustrating an example of a user authentication method, according to an embodiment. At step 102, a trained convolutional neural network (CNN) extracts one or more first features associated with a first eye pupil image of a user. At step 104, the same trained CNN extracts one or more second features associated with a second eye pupil image of the user. The second eye pupil image of the user is captured whilst a white light (flash light) is projected onto the user's face. On the other hand, the first eye pupil image of the user is captured without a white light (flash light) being projected onto the user's face.

[0040] The CNN can be trained using a large dataset of eye pupil images including eye pupils having the white light projected thereon, and also a large dataset of eye pupil images including eye pupils without having the white light projected thereon. The CNN can adopt a resnet18 network structure.

[0041] The source of the white flash light can be an in-built flash light of a smartphone with integrated camera. The smartphone's camera can be used to capture both the first eye pupil image of the user (i.e. without the in-built flash light activated) and the second eye pupil image of the user (i.e. whilst the in-built flash light is activated).

[0042] In particular, the method can include the step of capturing, using an image sensor (of the smartphone's camera), a first face image of the user including the first eye pupil image. Thereafter, an eye detection method is applied to the first face image of the user to generate a first eye bounding box. An area of the first face image of the user within the first eye bounding box corresponds to the first eye pupil image. Similarly, the method can include the step of capturing, using the image sensor, a second face image of the user including the second eye pupil image. Thereafter, the eye detection method is applied to the second face image of the user to generate a second eye bounding box. An area of the second face image of the user within the second eye bounding box corresponds to the second eye pupil image.

[0043] At step 106, a similarity score is generated based on the extracted first feature(s) and the extracted second feature(s).

[0044] At step 108, the user is positively authenticated on a condition that the similarity score indicates a difference between the extracted first feature(s) and the extracted second feature(s).

[0045] In the case of a photograph, screenshot, 2D or 3D mask spoof face, a white dot cannot be seen in and/or around the eye pupil area when a flash light is projected onto the eye pupil area.

[0046] On the other hand, in the case of a live face, when a flash light is projected onto the eye pupil area, a white dot can be seen in and/or around the eye pupil area. Figure 2a shows an example of an eye pupil area image 202 without a flash light projected onto the eye. For comparison, figures 2b, 2c and 2d show examples of an eye pupil area image (of a live face) with a flash light projected onto the eye. In Figure 2b, a white dot/spot 204 can be seen in the eye pupil area. In Figure 2c, two white dots/spots 206 can be seen in the eye pupil area. In Figure 2d, a white dot/spot 208 can be seen around the eye pupil area. Depending on the number of flash lights and the angle of the flash light(s) with respect to the eye pupil, the white dot(s)/spot(s) may appear at different locations in and/or around the eye pupil area.

[0047] The presence of white dot(s)/spot(s) in and/or around the eye pupil area of a live face when a flash light is projected onto the eye is due to the flash light shining/reflecting off the optic nerve. This happens when the flash light entering the eye at a certain angle is reflected, causing a white eye effect. This is also known as “optic disc reflection” or “white reflex”.

[0048] In addition to the white eye effect, it is also possible that red dot(s)/spot(s) appear in and/or around the eye pupil area of a live face when the flash light is projected onto the eye. This is known as the red eye effect wherein the flash light occurs too fast for the pupil to close, therefore much of the flash light passes into the eye through the pupil, reflects off the back of the eye (where there is an ample amount of blood) and out through the pupil. The camera records this reflected light.

[0049] A white eye effect or red eye effect causes a difference between the first eye pupil image of the user and the second eye pupil image of the user. Consequently, a difference exists between the extracted first features and the extracted second features. The user can be positively authenticated if the similarity score (determined at step 106) indicates a difference between the extracted first features and the extracted second features.

[0050] In an implementation, for more robust fraud detection and greater accuracy, the method can further include the following steps. First, a trained supplementary convolutional neural network (CNN) is used to extract one or more supplementary

features associated with a supplementary face image of the user. The supplementary face image of the user is captured whilst a non-white light is projected onto the user's face. Thereafter, a confidence score is generated based on the extracted supplementary feature(s).

[0051] In an implementation, the confidence score is predicted using a softmax function of a trained CNN. The softmax function is used in the final layer of a neural network-based classifier. The softmax function is used to map the non-normalized output of the network to a probability distribution over predicted output classes. The supplementary CNN is trained on large-scale face image sequences, wherein the face images are live face images and spoof face images. During the authentication phase, the confidence score is generated from the trained CNN with softmax function based on the captured face image(s) whilst non-white light(s) is/are projected onto the user's face.

[0052] For more robust fraud detection and greater accuracy, the user is positively authenticated on a condition that (i) the similarity score indicates the difference between the extracted first features and the extracted second features or (ii) the confidence score is greater than a pre-determined confidence threshold.

[0053] The supplementary CNN is different from the CNN described above at step 102. The supplementary CNN can be trained using a large dataset of face images including live face images and spoof face images having the non-white light projected thereon. The supplementary CNN can adopt a resnet18 network structure.

[0054] For even more robust fraud detection and even greater accuracy, 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, e.g. blue light, followed by red light, followed by yellow light, etc. The source of the non-white lights can be a display screen of a smartphone. The smartphone display screen can be configured to display a blue screen, followed by red screen, followed by yellow screen, etc. As each color is displayed, the image sensor (of the smartphone's camera) can be used to capture the user's face image. In other words, if three different non-white lights are displayed and consequently projected onto the user's face, three different face images of the user are captured.

[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.

[0061] In an alternative embodiment, the secondary memory 310 may additionally or alternatively include other similar devices for allowing computer programs or other instructions to be loaded into the computing device 300. Such devices can include, for example, a removable storage unit 322 and an interface 320. Examples of a removable storage unit 322 and interface 320 include a removable memory chip (such as an EPROM or PROM) and associated socket, and other removable storage units 322 and interfaces 320 which allow software and data to be transferred from the removable storage unit 322 to the computer system 300.

[0062] The computing device 300 also includes at least one communication interface 324. The communication interface 324 allows software and data to be transferred between computing device 300 and external devices via a communication path 326. In various embodiments, the communication interface 324 permits data to be transferred between the computing device 300 and a data communication network, such as a public data or private data communication network. The communication interface 324 may be used to exchange data between different computing devices 300 which such computing devices 300 form part an interconnected computer network. Examples of a communication interface 324 can include a modem, a network interface (such as an Ethernet card), a communication port, an antenna with associated circuitry and the like. The communication interface 324 may be wired or may be wireless. Software and data transferred via the communication interface 324 are in the form of signals which can be electronic, electromagnetic, optical or other signals capable of being received by communication interface 324. These signals are provided to the communication interface via the communication path 326.

[0063] Optionally, the computing device 300 further includes a display interface 302 which performs operations for rendering images to an associated display 330 and an audio interface 332 for performing operations for playing audio content via associated speaker(s) 334.

[0064] As used herein, the term "computer program product" may refer, in part, to removable storage unit 318, removable storage unit 322, a hard disk installed in hard disk drive 312, or a carrier wave carrying software over communication path 326 (wireless link or cable) to communication interface 324. Computer readable storage media refers to any non-transitory tangible storage medium that provides recorded instructions and/or data to the computing device 300 for execution and/or processing.

Examples of such storage media include floppy disks, magnetic tape, CD-ROM, DVD, Blu-ray™ Disc, a hard disk drive, a ROM or integrated circuit, USB memory, a magneto-optical disk, or a computer readable card such as a PCMCIA card and the like, whether or not such devices are internal or external of the computing device 300. Examples of transitory or non-tangible computer readable transmission media that may also participate in the provision of software, application programs, instructions and/or data to the computing device 300 include radio or infra-red transmission channels as well as a network connection to another computer or networked device, and the Internet or Intranets including e-mail transmissions and information recorded on Websites and the like.

[0065] The computer programs (also called computer program code) are stored in main memory 308 and/or secondary memory 310. Computer programs can also be received via the communication interface 324. Such computer programs, when executed, enable the computing device 300 to perform one or more features of embodiments discussed herein. In various embodiments, the computer programs, when executed, enable the processor 304 to perform features of the above-described embodiments. Accordingly, such computer programs represent controllers of the computer system 300.

[0066] Software may be stored in a computer program product and loaded into the computing device 300 using the removable storage drive 314, the hard disk drive 312, or the interface 320. Alternatively, the computer program product may be downloaded to the computer system 300 over the communications path 326. The software, when executed by the processor 304, causes the computing device 300 to perform functions of embodiments described herein.

[0067] It is to be understood that the embodiment of Figure 3 is presented merely by way of example. Therefore, in some embodiments one or more features of the computing device 300 may be omitted. Also, in some embodiments, one or more features of the computing device 300 may be combined together. Additionally, in some embodiments, one or more features of the computing device 300 may be split into one or more component parts.

[0068] Figure 4 is a schematic diagram illustrating an example of a user authentication system 400, according to an embodiment. The user authentication system comprises an extracting device 402, a score generating device 404 and an authenticating device

406. The extracting device 402 extracts, using a trained convolutional neural network (CNN), one or more first features associated with a first eye pupil image of a user. The extracting device 402 also extracts, using the trained CNN, one or more second features associated with a second eye pupil image of the user. The second eye pupil image of the user is captured whilst a white light is projected onto the user's face.

[0069] The score generating device 404 generates a similarity score based on the extracted first feature(s) and the extracted second feature(s). The authenticating device 406 positively authenticates the user on a condition that the similarity score indicates a difference between the extracted first feature(s) and the extracted second feature(s).

[0070] The extracting device 402 can also extract, using a trained supplementary convolutional neural network (CNN), one or more supplementary features associated with a supplementary face image of the user. 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 404 can generate a confidence score based on the extracted supplementary features. The authenticating device 406 positively authenticates the user on a condition that (i) the similarity score indicates the difference between the extracted first features and the extracted second features or (ii) the confidence score is greater than a pre-determined confidence threshold.

[0071] The system 400 can further include an image sensor 408 that captures a first face image of the user (including the first eye pupil image) and applies an eye detection method to the first face image of the user to generate a first eye bounding box. An area of the first face image of the user within the first eye bounding box corresponds to the first eye pupil image. The image sensor 408 also captures a second face image of the user (including the second eye pupil image) and applies the eye detection method to the second face image of the user to generate a second eye bounding box. An area of the second face image of the user within the second eye bounding box corresponds to the second eye pupil image.

[0072] The CNN can be trained using a large dataset of eye pupil images including eye pupils having the white light projected thereon. The supplementary CNN can be trained using a large dataset of face images including live face images and spoof face images having the non-white light projected thereon. Both the CNN and supplementary CNN can adopt a resnet18 network structure.

[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.