SYSTEM, METHOD AND COMPUTER READABLE MEDIUM FOR ALIGNING IDENTIFICATION DOCUMENT BASED ON IMAGE PROCESSING AND DEEP LEARNING TECHNIQUES

TECHNICAL FIELD

[0001] The present disclosure relates generally to a system, method and computer readable medium for aligning Identity (ID) documents. Embodiments of the disclosure are related to a system, method and computer readable medium for aligning ID documents in processes involving capturing the ID documents using hybrid methods, such as image processing and deep learning techniques.

BACKGROUND

[0002] In the present times, document scanning serves an integral function in every field and industry imaginable. With a plethora of tasks being performed virtually, the requirement of scanned documents has increased. Document scanning as a function is helpful for enhanced information preservation, enhanced information retrieval, better data security, improved collaboration within a workflow, improved disaster recovery, document verification, document analysis and the like. Usually, a flatten document image can be obtained by properly placing the document to be scanned on the glass window of a specialized flatbed scanner. However, as a reflection and consequence of the increased relevance and need for a more convenient option, nowadays people own personal digital devices such as digital cameras and smartphones to capture images at any time, said personal devices now being commonly used to scan documents.

[0003] While on one hand, conventional scanning devices allow a consistent and configurable image quality, said devices are usually bulky, inconvenient, expensive and constrained in terms of availability. On the other hand, conventional document scanning via personal devices offers a flexible and convenient way to have more options for document digitization, said option also suffers from several drawbacks such as the quality of the scanned image suffers from shadows, distortion, reflections, and blurriness depends on how it is captured, and the document distortion in digital images

is relatively critical and has a high impact on the processing, like document verification, document analysis and the like.

[0004] Document alignment is referred to as flattening the image to correct the distortion of the target. Alignment of a document entails the visual aspects of the documents, including but without limitation, paper size, margins, contents, and form design, and the alignment of the presentation of data. It is usually designed to shift the target to the center of the image and correct the orientation and size to best fit the image.

[0005] Conventional document alignment solutions allow users to manually select the vertices of the document in the image as reference points and the application completes the alignment afterwards based on the selected vertices. One disadvantage of said solutions is that they are limited to cases where the amount of images are small since they involve human intervention.

[0006] Other conventional document alignment solutions provide features to automatically correct the distortion. Such conventional solutions often use traditional image processing techniques, for example, by feature extraction such as Scale-Invariant Feature Transformation (SIFT) and Speeded-Up Robust Features (SURF) to detect a set of feature points followed by matching the feature points to the corresponding feature points on the flatten reference image, and perspective transformation to dewarp the image based on the mapping. A disadvantage of said solutions is that the quality of the input and reference image are expected to be similar so that the algorithm is able to match the two sets of feature points correctly. In addition, the feature point detection in such solutions works well mostly on high quality images with salient details such as clear shapes and textures. This evidently poses severe difficulty and inconvenience on the user as such conditions are often not available and/ or possible, thereby resulting in an unstable performance on lower quality images suffering from blurriness, reflection or insufficient brightness.

[0007] Other conventional solutions that also use traditional image processing involve detecting the boundary of the document together with the geometric property of the document content such as the orientation of the text. One drawback of said

solutions is that they are based on the assumption that the document is distinctive so the document can be separated from the background properly. Much like other conventional solutions, said solution also fails to take into account real-life scenarios wherein the document is not clearly distinguishable from the background.

[0008] Other conventional solutions include solutions employing deep learning techniques, involving single or a composite deep neural network, and Generative Adversarial Network (GAN) based solutions. One severe disadvantage of such solutions is the requirement for a large amount of training data as sufficient training data is necessary to build a robust neural network model with satisfying performance. Additionally, another disadvantage of said solutions is the requirement of higher computing power and reduced efficiency in terms of speed. Furthermore, the hardware specification is higher in those approaches since they include all the calculation in deep learning models.

[0009] Owing to the fact that obtaining a flatten image is critical in many applications such as Optical Character Recognition (OCR), facial recognition, and Virtual Reality (VR), there is a need for a document identity solution capable of overcoming the drawbacks of the conventional solutions.

SUMMARY

[0010] The present disclosure relates generally to a system, method and computer readable medium for aligning Identity (ID) documents. Embodiments of the disclosure are related to a system, method and computer readable medium for aligning ID documents in processes involving capturing the ID documents using hybrid methods, such as image processing and deep learning techniques.

[0011] The systems, methods and computer readable medium disclosed in the present application are advantageous in several ways, including but not limited to, providing a robust solution to document alignment of captured ID documents having different brightness, orientations and/ or resolutions, or being of a lower quality and suffering from blurriness, reflection and/ or insufficient brightness, reducing the training data required for segmentation, matching the feature points in a straightforward manner

without requiring any calculations, reducing the consumption of computing power, increasing the efficiency in terms of speed and accuracy of alignment, being capable of aptly and correctly aligning regions of ID documents with none or less landmarks, being capable of performing any tasks requiring image alignment, being capable of performing recognition tasks, such as facial recognition, overcoming the requirement of human intervention, saving costs, time, manual effort for aligning a large number of ID documents, overcoming requirements of higher hardware specifications as utilized in deep learning models.

[0012] In a first aspect of the present disclosure, a system for aligning at least one Identity (ID) document is disclosed. The system may comprise: a neural network model for obtaining one or more segmented landmarks in the ID document, wherein the ID document is captured at one or more angles, at one or more locations such that said one or more segmented landmarks in the ID document are obtained by performing a landmark segmentation process; and an image processing unit for processing the ID document by performing shape approximation on said one or more segmented landmarks, extracting a plurality of feature points of said one or more segmented landmarks, augmenting and dewarping the ID document by performing a perspective transformation of said plurality of feature points.

[0013] According to an embodiment in conjunction to the first aspect of the present disclosure, the image processing unit augments said plurality of feature points to cover at least one non-landmark area to obtain at least one augmented feature point.

[0014] According to an embodiment in conjunction to the first aspect of the present disclosure, in the perspective transformation, a projection matrix is computed using a plurality of reference points of said plurality of feature points and said at least one augmented feature point on a reference identity image and the identity document is dewarped by multiplying the projection matrix with the identity document.

[0015] According to an embodiment in conjunction to the first aspect of the present disclosure, the neural network model is a U-net model.

[0016] According to an embodiment in conjunction to the first aspect of the present disclosure, the neural network model is trained using at least one image of the ID document.

[0017] According to an embodiment in conjunction to the first aspect of the present disclosure, the neural network model comprises at least one encoder for extracting at least one image feature in said one or more segmented landmarks in the ID document.

[0018] According to an embodiment in conjunction to the first aspect of the present disclosure, the neural network model comprises at least one decoder for performing localization of said at least one image feature.

[0019] According to an embodiment in conjunction to the first aspect of the present disclosure, said plurality of feature points are extracted from a shape approximated landmark.

[0020] According to an embodiment in conjunction to the first aspect of the present disclosure, said at least one augmented feature point is obtained by calculating one or more interpolated points of said plurality of feature points in an upper and/ or lower region of the landmarks in the ID document, using a linear interpolation.

[0021] According to an embodiment in conjunction to the first aspect of the present disclosure, said at least one augmented feature point is obtained by calculating at least one centroid of said plurality of feature points of said one or more segmented landmarks.

[0022] According to an embodiment in conjunction to the first aspect of the present disclosure, said at least one augmented feature point is obtained by calculating an intersection point of the lines defined by said plurality of feature points of said one or more segmented landmarks.

[0023] According to an embodiment in conjunction to the first aspect of the present disclosure, said at least one augmented feature point comprises one or more interpolated points, the centroid and the intersection point.

[0024] In a second aspect of the present disclosure, a method for aligning at least one ID document is disclosed. The method may comprise obtaining, by a neural network model, at least one landmark of the ID document at one or more angles, at one or more locations; performing, by the neural network model, at least one segmentation of said at least one landmark in the ID document to obtain one or more segmented landmarks; performing, by the image processing unit, at least one shape approximation on said one or more segmented landmarks to obtain at least one shape approximated landmark; extracting, by the image processing unit, a plurality of feature points from said one or more segmented landmarks and dewarping, by the image processing unit, the ID document using perspective transformation of said plurality of feature points.

[0025] According to an embodiment in conjunction to the second aspect of the present disclosure, the method further comprises: extracting, by the image processing unit, at least one augmented feature point from said plurality of feature points to cover at least one non-landmark area; and dewarping, by the image processing unit, the identity document using perspective transformation of said plurality of feature points and said at least one augmented feature point.

[0026] According to an embodiment in conjunction to the second aspect of the present disclosure, said one or more segmented landmarks is obtained by: extracting at least one image feature in said at least one landmark in the ID document using an encoder; performing localization of said at least one image feature using a decoder; and capturing and labeling at least one image feature as said one or more segmented landmarks.

[0027] According to an embodiment in conjunction to the second aspect of the present disclosure, said plurality of feature points are obtained by: performing, by the image processing unit, at least one rectangular approximation on said one or more segmented landmarks; obtaining said at least one shape approximated landmark; calculating a midpoint on each edge of said at least one shape approximated landmark; forming an axis along two opposite midpoints of the edges; obtaining a centroid at an intersection feature point of the axis; and obtaining said plurality of

feature points of the one or more segmented landmarks comprising the midpoints and the centroid.

[0028] According to an embodiment in conjunction to the second aspect of the present disclosure, said plurality of feature points are obtained by: performing, by the image processing unit, at least one elliptical approximation on said one or more segmented landmarks; obtaining said at least one shape approximated landmark; forming a major and a minor axis of the ellipse; obtaining the midpoint on the ellipse using intersection of the major and minor axes and a contour of the ellipse; obtaining the centroid of the ellipse using the midpoints on the ellipse; and obtaining said plurality of feature points of the one or more segmented landmarks comprising the midpoints and the centroid.

[0029] According to an embodiment in conjunction to the second aspect of the present disclosure, said at least one augmented feature point is obtained by calculating one or more interpolated points of said plurality of feature points in an upper and/or lower region of the landmarks in the ID document, using a linear interpolation.

[0030] According to an embodiment in conjunction to the second aspect of the present disclosure, said at least one augmented feature point is obtained by: calculating at least one centroid of said plurality of feature points of said one or more segmented landmarks; and calculating an intersection point of lines defined by said plurality of feature points of said one or more segmented landmarks.

[0031] According to an embodiment in conjunction to the second aspect of the present disclosure, said at least one augmented feature point comprises said one or more interpolated points, the centroid and the intersection point.

[0032] According to an embodiment in conjunction to the second aspect of the present disclosure, dewarping the ID document using perspective transformation of said plurality of feature points and said at least one augmented feature point comprises the steps of: obtaining a plurality of reference points of a reference image of the ID document, corresponding to said plurality of feature points and said at least one

augmented feature point; computing a projection matrix using the series of said plurality of reference points and said plurality of feature points and said at least one augmented feature point; and obtaining a dewarped image of the ID document by multiplying the projection matrix with the ID document.

[0033] In a third aspect of the present disclosure, a computer readable medium storing a program comprising instructions that, when executed by a system, cause the system to perform the method for aligning at least one ID document using hybrid methods, such as image processing and deep learning techniques is disclosed.

[0034] The following description is provided to facilitate an understanding of some of the innovative features unique to the disclosed embodiment and is not intended to be a full description. A full appreciation of the various aspects of the embodiments disclosed herein can be gained by taking into consideration the entire specification, claims, drawings, and abstract as a whole.

BRIEF DESCRIPTION OF THE DRAWINGS

[0035] **FIG. 1A** is a block diagram of an example system **100** for aligning an Identity (ID) document **102**, in accordance with the disclosed embodiments.

[0036] **FIG. 1B** is a block diagram of an example neural network model **104** for obtaining a segmented landmark of the ID document, in accordance with the disclosed embodiments.

[0037] **FIG. 2A, FIG. 2B, FIG. 2C, FIG. 2D** and **FIG. 2E** are diagrammatic representations of an example system flow **200** for alignment of an ID document **202** having a background **203** in an image **201**, in accordance with the disclosed embodiments.

[0038] **FIG. 3A**, **FIG. 3B** and **FIG. 3C** are diagrammatic representations of example images **302**, **304, 306** disclosing a semantic segmentation approach on different perspective views of a captured ID document **308** having a background, in accordance with the disclosed embodiments.

[0039] **FIG. 4A** and **FIG. 4B** are diagrammatic representations of an example segmented landmark **404** having a shape similar to a rectangle, in accordance with the disclosed embodiments.

[0040] **FIG. 5A** and **FIG. 5B** are diagrammatic representations of an example segmented landmark **504** having a shape similar to an ellipse, in accordance with the disclosed embodiments.

[0041] **FIG. 6A**, **FIG. 6B** and **FIG. 6C** are a diagrammatic representation of an example image **610**, **620** and **630** disclosing extraction of one or more augmented feature points, in accordance with the disclosed embodiments.

[0042] **FIG. 7A** is a diagrammatic representation of an example reference image **710** of an ID document. **FIG. 7B** is a diagrammatic representation of the reference image **710** disclosing a plurality of reference points corresponding to the extracted existing feature points and the augmented feature points, in accordance with the disclosed embodiments.

[0043] **FIG. 8A** is a diagrammatic representation of an example image **810** of an ID document used for perspective transformation process and **FIG. 8B** is a diagrammatic representation of example image **820** disclosing augmented feature points $\mathbf{p_1}$, $\mathbf{p_2}$, $\mathbf{p_3}$, $\mathbf{p_4}$, $\mathbf{p_5}$, $\mathbf{p_6}$, $\mathbf{m_1}$, $\mathbf{m_2}$, $\mathbf{g_1}$, $\mathbf{g_2}$ of the image **820** of the ID document, in accordance with the disclosed embodiments.

[0044] **FIG. 9A** is a diagrammatic representation of an example image **910** of an ID document that is captured or scanned and **FIG. 9B** is a diagrammatic representation of an example dewarped image **920** of the aligned ID document, in accordance with the disclosed embodiments.

[0045] **FIG. 10** is a flow diagram that illustrates an example method **1000** for aligning an ID document, in accordance with the disclosed embodiments.

[0046] **FIG. 11** is a flow diagram that illustrates an example method **1100** for

segmenting landmarks in an ID document, in accordance with the disclosed embodiments.

[0047] **FIG. 12** is a flow diagram that illustrates an example method **1200** for extracting feature points of a segmented landmark with a shape similar to a rectangular or a parallelogram on an ID document, in accordance with the disclosed embodiments.

[0048] **FIG. 13** is a flow diagram that illustrates an example method **1300** for extracting feature points of a segmented landmark with a shape similar to an ellipse on an ID document, in accordance with the disclosed embodiments.

[0049] **FIG. 14** is a flow diagram that illustrates an example method **1400** for obtaining augmented feature points, in accordance with the disclosed embodiments.

[0050] **FIG. 15** is a flow diagram that illustrates an example method **1500** for obtaining a dewarped image of an ID document using perspective transformation, in accordance with the disclosed embodiments.

[0051] The summary above, as well as the following detailed description of illustrative embodiments, is better understood when read in conjunction with the appended drawings. For the purpose of illustrating the present disclosure, exemplary constructions of the disclosure are shown in the drawings. However, the disclosure is not limited to specific methods and instrumentalities disclosed herein. Moreover, those in the art will understand that the drawings are not to scale. Wherever possible, like elements have been indicated by identical numbers.

DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

[0052] The particular configurations discussed in the following description are non-limiting examples that can be varied and are cited merely to illustrate at least one embodiment and are not intended to limit the scope thereof.

[0053] The present invention provides a system, method and computer readable medium for aligning Identity (ID) documents. ID documents may be captured for safekeeping, data security purposes, in document identity processes like verification of ID and other ID recognition tasks, collaboration or workflow purposes and/ or the like. Embodiments of the disclosure utilize deep learning based models with traditional image processing methods thereby resulting in a system, method and computer readable medium for aligning captured ID documents

[0054] The present disclosure utilizes a semantic segmentation approach to localize the landmarks on an ID document. Semantic segmentation approach is a pixel level classification which links each pixel in an image to a label of what is being represented. The present disclosure further involves ascertaining a combination of feature points of existing landmarks on the ID document as well as augmented feature points on the ID document. In computer vision and image processing, a feature is a piece of information about the content of an image; typically about whether a certain region of the image has certain properties. The present disclosure further involves dewarping the ID document using perspective transformation. Perspective transformation is used to project an image onto a new plane.

[0055] Various non-limiting benefits of the present disclosure includes, providing a robust solution to document alignment of captured ID documents having different brightness, orientations and/ or resolutions, or being of a lower quality and suffering from blurriness, reflection and/ or insufficient brightness. The present disclosure enables identity of documents by utilizing a relatively small neural network capable of being trained with a smaller training set, thereby saving costs, time, manual effort and other resources employed at the stage of pre-processing data. The present disclosure allows identity of documents in an efficient and accurate manner including regions of ID documents with reduced or none outstanding features. The matching model in the present disclosure is capable of matching the feature points in a straightforward manner without requiring any calculations. Further, the present disclosure utilizes perspective transformation done by traditional image processing methods thereby reducing the consumption of computing power and increasing efficiency in terms of speed. The present disclosure is advantageous over conventional solutions as it also overcomes the requirement of human intervention, saving costs, time, manual effort

for aligning a large number of ID documents.

[0056] The alignment solution of the present disclosure is capable of being adapted for any recognition tasks such as facial recognition requiring image alignment for subsequent processing. In some embodiments of the present invention, the disclosure can be adapted to segment the landmarks such as eyes, noses, and mouth and augment the feature points to cover the entire face. The frontal face can be obtained by the perspective transformation based on the feature points.

[0057] **FIG. 1A** is a block diagram of an example system **100** for aligning an Identity (ID) document **102** (also referred as identity document **102**), in accordance with the disclosed embodiments. The system **100** comprises a neural network model **104.** In an embodiment of the present invention, one or more segmented landmarks in the ID document are captured and labeled by the neural network model **104**. The ID documents may be captured at one or more angles, at one or more locations and one or more segmented landmarks in the ID document are obtained by performing a landmark segmentation process. The system **100** also comprises an image processing unit **106**. In an embodiment of the present invention, the image processing unit **106** processes the ID document **102** by performing shape approximation on the segmented landmarks, and extracts plurality of feature points of at least one landmark of the ID document **102**, and augments the feature points to cover at least one non-landmark areas to extract at least one augmented feature point and dewarps the ID document **102** using perspective transformation process. In some embodiments of the present invention, when the feature points sufficiently cover all the areas of the ID document, perspective transformation process may be performed using only said feature points. In some embodiments of the present invention, when the feature points partially cover the areas of the ID document, perspective transformation process may be performed using the feature points and augmented feature points.

[0058] The output of the image processing unit **106** is an aligned ID **108**. To perform one or more of the perspective transformation process, feature point extraction, shape approximation and feature point augmentation, the present invention, the image processing unit **106** such as a traditional image processing unit is utilized, whereas the process landmark segmentation is performed by the neural network model **104**,

such as a U-Net model**.** Thus, present invention is a hybrid processing approach using both the neural network model **104** and the image processing unit **106**. The present invention can be deployed in a computing device, not limited to mobile or can use cloud Application Programming Interface (API) without limitation, to flatten the ID document API at the early stage, so that more information can be exposed in the subsequent recognition steps.

[0059] It should be noted that the image processing unit **106** comprises, without limitation, one or more algorithms, optimization solutions and functions to perform the invention process such as, performing shape approximation on the segmented landmarks, extracting plurality of feature points of at least one landmark of the ID document **102**, augmenting the feature points to cover at least one non-landmark areas to obtain at least one augmented feature points and dewarping the ID document using a perspective transformation.

[0060] For each landmark, at least one feature point is extracted based on the approximated shape. Additional feature points can be obtained by augmenting the feature points based on the approximated landmark shape to cover non-landmark areas. Obtaining augmented feature points is not mandatory, but the same may increase the warping accuracy especially for IDs with less landmarks and may be preferred in such scenarios. Accordingly, it is to be noted that higher coverage of feature points throughout the ID document boosts the perspective transformation accuracy and the same may be achieved by augmenting such feature points, as disclosed in the present invention.

[0061] **FIG. 1B** is a block diagram of an example neural network model **104** for obtaining a segmented landmark of the ID document, in accordance with the disclosed embodiments. The example neural network model **104** disclosed in **FIG. 1B** is a U-net model. In some embodiments of the present invention, said neural networks are convolutional neural networks such as U-net. U-net is one of the well-known and commonly employed convolutional neural networks being a Fully Convolutional Network (FCN) that does image segmentation and is referred so due to its symmetric shape, which is different from other FCN variants. U-net is an end-to-end fully-

convolutional network containing a contraction part (encoder) to extract the image features followed by an expansion path (decoder) for object localization. Convolutional neural networks such as U-net are preferred as they don't involve any dense layer in the architecture, thereby reducing the number of parameters and computation time. In an embodiment of the present invention, the neural network model **104** comprises at least one encoder **114** and at least one decoder **116**. The encoder **114** and the decoder **116** are used in the process of landmark segmentation of the ID document. The encoder **114** is used for extracting at least one image feature in each landmark in the ID document. The decoder **116** is used for performing localization of the image feature. In general, the encoder **114** and the decoder **116** consist of several neural network layers and are used in processing of the landmark segmentation and the feature extraction process.

[0062] **FIG. 2A, FIG. 2B, FIG. 2C, FIG. 2D** and **FIG. 2E** are diagrammatic representations of an example system flow **200** for alignment of an ID document **202** having a background **203** in an image **201**, in accordance with the disclosed embodiments. The present invention combines the traditional image processing method with deep learning based models and makes use of the advantages of the two schemes. Embodiments of the present invention utilize a semantic segmentation approach to localize the landmarks on an ID document, that is, extract the contours of the landmarks. Semantic segmentation approach is a pixel level classification which links each pixel in an image to a label of what is being represented. Embodiments of the present invention employ a neural network model for semantic segmentation to localize and extract the border of each landmark on the ID document, such as national flag, chip, photos and/or the like.

[0063] As disclosed in **FIG. 2B**, the system flow **200** involves capturing and labeling each significant landmark **204**, **205**, **206**, **207** on the ID document as an isolated region. The neural network model in the embodiments of the present invention is trained with images captured from different angles and covering different artifacts including blurriness, reflections or partial occlusion for each type of ID. For example, the U-net model may be trained with 2300 image pairs, each pair consists of an original image and a corresponding labeled image. In the labeled image, the pixel value is associated with the class the pixel belongs to, for example, the pixels forming the

national flag on the ID are labeled together.

[0064] The system flow **200** further involves ascertaining feature points **208**, **209**, **210**, **211** of existing landmarks **204**, **205**, **206**, **207** on the ID document, as disclosed in **FIG. 2C**. The feature points are extracted within the contours of the landmarks based on the shapes of said landmarks. In some embodiments of the present invention, the shapes of the landmarks on an ID document may be the same, or may be varied in some other embodiments, such as rectangle or ellipse. In some embodiments of the present invention, the contours of a landmark that is similar to a rectangle or parallelogram may be approximated to a rectangle and the midpoint on each edge may be calculated. The four vertices where the two axes connecting the four midpoints intersect the contour and the centroid of the contour may be taken as the feature point(s). In some embodiments of the present invention, the contours of a landmark that is similar to an ellipse or a circle may be approximated to an ellipse and the major and minor axes may be calculated.

[0065] The system flow **200** further involves augmenting feature points **212** based on the feature points of the existing landmarks on the ID document, as disclosed in **FIG. 2D**. In some embodiments of the present invention, feature points may be augmented to cover regions of the ID document where there is a lack or absence of existing landmarks. An advantage of the embodiments of the present invention is that the augmented feature points significantly enhance and improve the accuracy of perspective transformation by increasing the number of reliable feature points and the coverage of the targets on the ID document. Thereby, another advantage of the embodiments of the present invention is that superior dewarped results can be expected on account of introduction of more feature points located at different parts of the target. Embodiments of the present invention may involve any method for augmenting the feature points. In some embodiments of the present invention, a simple linear algebra concept may be applied for augmenting feature points. In some embodiments of the present invention, advanced methods such as affine transformation may be applied in feature point augmentation.

[0066] The system flow **200** further involves dewarping the ID document using perspective transformation. Perspective transformation is used to project an image

onto a new plane. For each ID document, a reference image *R* with a plurality of reference points corresponding to the feature points is prepared. The projection matrix *H* is computed from formula ***P' = HP*** where *P* is the extracted feature point combination including the feature points from the landmarks and augmented points and *P'* is the corresponding feature points in *R*. The final dewarped image *I'* **214** as disclosed in **FIG. 2E** can be obtained by multiplying the projection matrix H by the image I. The region wherein the ID document is mapped is cropped to result in the final dewarped image *I'* **214**. The cropped region is the final dewarped image having the same size with the reference image R.

[0067] The traditional image processing methods are used to perform the steps, not limited to, the shape approximation from the segmented contours, feature point extraction, feature point augmentation and the perspective transformation. Using the traditional image processing methods, the landmark segmentation image from the neural network model such as the U-Net model is converted to an aligned image.

[0068] **FIG. 3A**, **FIG. 3B** and **FIG. 3C** are diagrammatic representations of example images **302**, **304, 306** disclosing a semantic segmentation approach on different perspective views of a captured ID document **308**, in accordance with the disclosed embodiments. Images **302**, **304, 306** represent the images of ID document **308** taken at different angles and locations. As disclosed in **FIG. 3A**, **FIG. 3B** and **FIG. 3C**, the significant landmarks **310, 312, 314, 316** are captured and labeled on the ID document **308** as an isolated region and the contours of the landmarks **310, 312, 314, 316** are outlined in **FIG. 3A**, **FIG. 3B** and **FIG. 3C**.

[0069] Once the landmarks in the ID document are localized the present invention performs the process such as, shape approximation of landmarks and extracting feature points from each landmark. **FIG. 4A** and **FIG. 4B** are diagrammatic representations of an example segmented landmark **404** having a shape similar to a rectangle, in accordance with the disclosed embodiments. The segmented landmark **404** is approximated to a rectangle **402** using a process of a rectangular shape approximation by the neural network model, for example U-net model, as shown in **FIG. 4A**. As disclosed in **FIG. 4B**, the midpoints on each edge are calculated and the four vertices where the two axes connecting the four midpoints intersect the contour

p_1, p_2, p_3, p_4 and the centroid of the contour **p_5** may be taken as the feature points.

[0070] **FIG. 5A** and **FIG. 5B** are diagrammatic representations of an example segmented landmark **504** having a shape similar to an ellipse, in accordance with the disclosed embodiments. The segmented landmark **504** is approximated to a circular or an elliptical shape **502** using a process of an elliptical shape approximation by the neural network model, for example U-net model, as shown in **FIG. 5A.** As disclosed in **FIG. 5B**, the major and minor axes are calculated and the centroid of the contour **p_5** and four intersection points **p_1, p_2, p_3, p_4** of the two aces and the contour are extracted as feature points.

[0071] In general, shape approximation is a classical mathematical problem done by an optimization solution which estimates the parameters of the shape, such as, rectangle or ellipse, by minimizing the distance between the scattered points and shape. In some embodiments of the present invention, all the common geometric shapes can be used as the approximation of the contours as long as the feature points of the shape can be well defined. Rotated rectangle and ellipses are the most commonly used landmarks on the ID documents and accordingly exemplified without limitation in the above disclosure.

[0072] **FIG. 6A**, **FIG. 6B** and **FIG. 6C** are a diagrammatic representation of an example image **610**, **620** and **630** disclosing extraction of one or more augmented feature points, in accordance with the disclosed embodiments. In some embodiments of the present invention, the augmented feature points cover regions of the ID document where there is a lack or absence of existing or outstanding landmarks. An advantage of the embodiments of the present invention is that the augmented feature points significantly enhance and improve the accuracy of perspective transformation by increasing the number of reliable feature points and the coverage of the targets on the ID document. Thereby, another advantage of the embodiments of the present invention is that superior dewarped results can be expected on account of introduction of more feature points located at different parts of the target. Embodiments of the present invention extract the augmented feature points covering more regions based on the feature points of the landmarks obtained using the shape approximation and feature extraction process depicted in **FIG. 4A, FIG. 4B, FIG. 5A** and **FIG. 5B**.

[0073] In some embodiments of the present invention, a simple linear algebra concept may be applied for augmenting feature points. In some embodiments of the present invention, advanced methods such as affine transformation may be applied in feature point augmentation. In some embodiments of the present invention, augmented feature points may be extracted by calculating at least one of points between two existing feature points, calculating a centroid of multiple points or calculating intersection of two lines formed of existing feature points.

[0074] The augmented feature points between two existing feature points that contribute to the feature point coverage can be easily calculated, for example, by linear interpolation. For a given two points p_1 and p_2, the interpolated points between p_1 and p_2 can be represented as $p_k = a*p_1 + b*p_2$ where a and b in [0, 1] and a + b =1.

[0075] **FIG. 6A** discloses four segmented landmarks, namely **602**, **604**, **606** and **608** having corresponding feature points $\mathbf{p_1}$, $\mathbf{p_4}$, $\mathbf{p_2}$ and $\mathbf{p_3}$ respectively. In embodiments of the present invention, semantic segmentation approach may be utilized to localize the landmarks on an ID document, that is, extract the contours of the landmarks. Embodiments of the present invention employ a neural network model for semantic segmentation to localize and extract the border of each landmark on the ID document, such as national flag, chip, photos and/or the like. In some embodiments of the present invention, said neural networks are convolutional neural networks such as U-net. In embodiments of the present invention, feature points $\mathbf{p_1}$, $\mathbf{p_4}$, $\mathbf{p_2}$ and $\mathbf{p_3}$ may be extracted or ascertained using shape approximation and feature extraction process. As disclosed in **FIG. 6A**, in some embodiments of the present invention, the augmented feature points or the interpolated points $\mathbf{m_1}$ and $\mathbf{m_2}$ may be extracted using linear interpolation of the feature points $\mathbf{p_1}$ and $\mathbf{p_4}$. The interpolated points $\mathbf{m_3}$, $\mathbf{m_4}$ and $\mathbf{m_5}$, also referred to as augmented feature points $\mathbf{m_3}$, $\mathbf{m_4}$ and $\mathbf{m_5}$, are obtained using linear interpolation of the feature points $\mathbf{p_2}$ and $\mathbf{p_3}$. The interpolated points $\mathbf{m_1}$ and $\mathbf{m_2}$ cover the upper center region of the ID document and $\mathbf{m_3}$, $\mathbf{m_4}$, $\mathbf{m_5}$ increase the coverage in the bottom region of the ID document. An advantage of using linear interpolation for extraction of augmented feature points is that it works as a complement of interpolation to expand the feature points toward the card holder, the same being specifically helpful in cases where landmarks are located relatively far

from the border of the IDs.

[0076] In some embodiments of the present invention, the augmented feature points may be extracted using the centroid of multiple points, as disclosed in **FIG. 6B. FIG. 6B** discloses four segmented landmarks, namely **602**, **604**, **606** and **608** having corresponding feature points $\mathbf{p_1}$, $\mathbf{p_4}$, $\mathbf{p_2}$ and $\mathbf{p_3}$ respectively. As disclosed in **FIG. 6B**, the augmented feature point(s) or the interpolated point(s) $\mathbf{g_1}$ may be extracted by calculating the centroid or weighted average of multiple known or existing feature points $\mathbf{p_1}$, $\mathbf{p_4}$, $\mathbf{p_2}$. The feature points of one or more nearby landmarks and feature points of one or more separated landmarks may be identified. The centroid $\mathbf{g_1}$ is also referred to as the centroid and it corresponds to the augmented feature point or augmented point.

[0077] In some embodiments of the present invention, the augmented feature points may be extracted using intersection points of the lines defined by any two or more known or existing feature points, thereby making the augmented set more comprehensive. **FIG. 6C** discloses four segmented landmarks, namely **602**, **604**, **606** and **608** having corresponding feature points $\mathbf{p_1}$, $\mathbf{p_4}$, $\mathbf{p_2}$ and $\mathbf{p_3}$ respectively. As disclosed in **FIG. 6C**, the augmented feature point $\mathbf{I_1}$ is the intersection of two lines defined by the feature points $\mathbf{p_1}$, $\mathbf{p_2}$, $\mathbf{p_3}$, $\mathbf{p_4}$ that reinforces the center region of the ID document.

[0078] It should be noted that the one or more feature point augmentation methods shown in **FIG. 6A**, **FIG. 6B** and **FIG. 6C** are used without limitation depending on the ID layouts. The feature point augmentation methods are used to spread the feature points evenly and comprehensively on the ID document so that the entire document can be covered. As disclosed above, an advantage of the above-discussed feature point augmentation methods is that regions of the ID document that lack outstanding landmarks are also clearly and efficiently aligned. Further, another advantage of said feature point augmentation methods is that while pursuing matching and dewarping processes, embodiments of the present invention are able to efficiently match landmarks and their corresponding positions as feature points are extracted from each landmark and additionally, augmented feature points corresponding to said existing feature points are also extracted.

[0079] The approaches in figures **FIG. 6A**, **FIG. 6B** and **FIG. 6C** are three exemplary ways based on interpolation, other approaches such as extrapolation can be applied wherever applicable.

[0080] **FIG. 7A** is a diagrammatic representation of an example reference image **710** of an ID document. **FIG. 7B** is a diagrammatic representation of the reference image **710** disclosing a plurality of reference points corresponding to the extracted existing feature points and the augmented feature points, in accordance with the disclosed embodiments. Embodiments of the present invention may extract the augmented feature points covering more regions based on the feature points of the landmarks obtained using the shape approximation and feature extraction process depicted in **FIG. 4A, FIG. 4B, FIG. 5A** and **FIG. 5B**. In some embodiments of the present invention, the augmented feature points may be extracted using the exemplary interpolation techniques disclosed in **FIG. 6A**, **FIG. 6B** and **FIG. 6C** or by other techniques such as extrapolation or affine transformation. For each type of ID document, a reference image *R* **710** with a plurality of reference points $\mathbf{p_1'}$, $\mathbf{p_2'}$, $\mathbf{p_3'}$, $\mathbf{p_4'}$, $\mathbf{p_5'}$, $\mathbf{p_6'}$ $\mathbf{m_1'}$, $\mathbf{m_2'}$, $\mathbf{g_1'}$, $\mathbf{g_2'}$ corresponding to the feature points of ID document to be aligned is prepared.

[0081] The projection matrix *H* is computed from formula $P' = HP$, where *P* is equal to $\{p_1, p_2, \ldots, p_k\} + \{g_1, g_2, \ldots, g_l\} + \{m_1, m_2, \ldots, m_n\}$. *P* is the extracted feature point combination including the feature points from the landmarks and augmented points as shown in **FIG. 6A**, **FIG. 6B** and **FIG. 6C**. The numbers k, l and n represent the number of feature points in p, g and m respectively. The numbers k, l and n will depend on the ID document and can be extracted using figures **FIG. 4A, FIG. 4B, FIG. 5A**, **FIG. 5B, FIG. 6A**, **FIG. 6B** and **FIG. 6C**.

[0082] *P'* is equal to $\{p_1', p_2', \ldots, p_k'\} + \{g_1', g_2', \ldots, g_l'\} + \{m_1', m_2', \ldots, m_n'\}$ and refers to the corresponding feature points in reference image *R* displayed as shown in **FIG. 8B**. The numbers k, l, n are the amount of each set. The numbers k, l and n represent the number of reference points in **p'**, **g'** and **m'** respectively. The numbers k, l and n will depend on the reference document and can be extracted using figure **FIG. 7C**.

[0083] **FIG. 8A** is a diagrammatic representation of an example image **810** of an ID document used for perspective transformation process and **FIG. 8B** is a diagrammatic representation of example image **820** disclosing augmented feature points $\mathbf{p_1}$, $\mathbf{p_2}$, $\mathbf{p_3}$, $\mathbf{p_4}$, $\mathbf{p_5}$, $\mathbf{p_6}$, $\mathbf{m_1}$, $\mathbf{m_2}$, $\mathbf{g_1}$, $\mathbf{g_2}$ of the image **820** of the ID document, in accordance with the disclosed embodiments. The ID card using perspective transformation. Perspective transformation is used to project an image onto a new plane.

[0084] **FIG. 9A** is a diagrammatic representation of an example image **910** of an ID document that is captured or scanned and **FIG. 9B** is a diagrammatic representation of an example dewarped image **920** of the aligned ID document, in accordance with the disclosed embodiments. Embodiments of the present invention dewarp the captured image **910** using perspective transformation process. The final dewarped image *I'* **920** can be obtained by multiplying the projection matrix *H* by the image *I*. The equation for obtaining the final dewarped image *I' (I'= HI)* is equal to the product of the projection matrix *H* and the image *I*. The system crops out the region where the ID is mapped to in *I'*. The cropped region is the final dewarped image *I'* **920** having the same size as the reference image *R* disclosed within the image **910** of **FIG. 9A**.

[0085] Reference image is obtained from a real flatten ID with standard layout. The feature points are selected from landmarks on the ID. The location of these reference points p_1', p_2', p_3', … p_k' are collected as the target points P'. For any new image, the feature points p_1, p_2, p_3, ... p_k at corresponding locations are collected using a neural network model, for example a U-Net model, and mathematical augmentation is performed. The locations of p_1, p_2, p_3, ... p_k are denoted set *P*. The projection matrix *H* shows the way to convert *P* to *P'*. Further, the reference image *R* provides the information pertaining to the placement and positioning of the significant points in a successfully aligned ID document. In the present disclosure, extraction of feature points of each landmark is conducted along with extraction of corresponding augmented feature points, thereby eliminating discrepancies in alignment on account of miscellaneous and external factors such as pixel information, like the color or texture of the ID document. Accordingly, embodiments of the present invention provide a robust solution to document alignment of captured ID documents having different brightness, orientations and/ or resolutions, or being of a lower quality and suffering from blurriness, reflection and/ or insufficient brightness. Further by eliminating any

dependence of the extraction of feature points in the reference images on other variant factors such as layout of the ID documents, embodiments of the present invention are able to efficiently and aptly align ID documents, thereby expanding the scope of application of the present invention. For the above reasons, the alignment solution of the present disclosure is capable of being adapted for any recognition tasks such as facial recognition requiring image alignment for subsequent processing. In some embodiments of the present invention, the disclosure can be adapted to segment the landmarks such as eyes, noses, and mouth and augment the feature points to cover the entire face. The frontal face can be obtained by the perspective transformation based on the feature points.

[0086] Any combination of feature points can be the reference sets for alignment as long as the way to extract the features points are the same in the new image and the reference image. For example, if p_1 in the new images is chosen as one of the feature points, then the corresponding p_1' in reference image should also be chosen. The feature points in the new image and reference image are mapped one to one. In some embodiments of the present invention, when the feature points sufficiently cover all the areas of the ID document, perspective transformation may be performed using only said feature points. In some embodiments of the present invention, when the feature points partially cover the areas of the ID document, perspective transformation may be performed using the feature points and augmented feature points.

[0087] **FIG. 10** is a flow diagram that illustrates an example method **1000** for aligning an ID document, in accordance with the disclosed embodiments. As disclosed in **FIG. 10**, the example method **1000** can be implemented by obtaining at least one landmark of the ID document at one or more angles and at one or more locations, as at block **1010**. Thereafter, as at block **1020**, at least one segmentation of the at least one landmark in the ID document is performed to obtain at least one segmented landmark. Embodiments of the present invention employ a neural network model for semantic segmentation to localize and extract the border of each landmark on the ID document, such as national flag, chip, photos and/or the like. In some embodiments of the present invention, said neural networks are convolutional neural networks such as U-net. In some embodiments of the present invention, each significant landmark is captured and labeled on the ID document as an isolated region. The neural network model in

the embodiments of the present invention is trained with images captured from different angles and covering different artifacts including blurriness, reflections or partial occlusion for each type of ID. For example, the U-net model may be trained with 2300 image pairs, each pair consists of an original image and a corresponding labeled image. In the labeled image, the pixel value is associated with the class the pixel belongs to, for example, the pixels forming the national flag on the ID are labeled together.

[0088] As at block **1030**, shape approximation is performed on the at least one segmented landmark to obtain at least one shape approximated landmark. Thereafter, as disclosed at block **1040**, plurality of feature points is extracted from the shape approximated landmarks. In some embodiments of the present invention, the shapes of the landmarks on an ID document may be the same, or may be varied in some other embodiments, such as rectangle or ellipse. In some embodiments of the present invention, the contours of a landmark that is similar to a rectangle or parallelogram may be approximated to a rectangle and the midpoint on each edge may be calculated. The four vertices where the two axes connecting the four midpoints intersect the contour and the centroid of the contour may be taken as the feature point(s). In some embodiments of the present invention, the contours of a landmark that is similar to an ellipse or a circle may be approximated to an ellipse and the major and minor axes may be calculated.

[0089] In some embodiments of the present invention, the method **1000** may be implemented by, for example, extracting at least one augmented feature point as at block **1050**. In some embodiments of the present invention, extracting augmented feature points may increase the warping accuracy especially for IDs with less landmarks and may be preferred in such scenarios. It is to be noted that higher coverage of feature points throughout the ID document boosts the perspective transformation accuracy and the same may be achieved by augmenting such feature points, as disclosed in the present invention. The augmented feature points are based on the plurality of feature points of the at least one landmark on the ID document. In some embodiments of the present invention, a simple linear algebra concept may be applied for augmenting feature points. In some embodiments of the present invention, advanced methods such as affine transformation may be applied in feature point

augmentation. In some embodiments of the present invention, augmented feature points may be extracted by calculating at least one of points between two existing feature points, calculating a centroid of multiple points or calculating intersection of two lines formed of existing feature points.

[0090] In some embodiments of the present invention, the augmented feature points between two existing feature points that contribute to the feature point coverage can be easily calculated, for example, by linear interpolation. For a given two points p_1 and p_2, the interpolated points between p_1 and p_2 can be represented as $p_k = a*p_1 + b*p_2$ where a and b in [0, 1] and a + b =1. In some embodiments of the present invention, the augmented feature points or the interpolated points may be extracted using linear interpolation of the feature points. The interpolated points, also referred to as augmented feature points**,** are obtained using linear interpolation of the feature points. The interpolated points cover the upper center region of the ID document and increase the coverage in the bottom region of the ID document. An advantage of using linear interpolation for extraction of augmented feature points is that it works as a complement of interpolation to expand the feature points toward the card holder, the same being specifically helpful in cases where landmarks are located relatively far from the border of the IDs. In some embodiments of the present invention, the augmented feature point(s) or the interpolated point(s) may be extracted by calculating the centroid or weighted average of multiple known or existing feature points. The feature points of one or more nearby landmarks and feature points of one or more separated landmarks may be identified. The centroid is also referred to as the centroid and it corresponds to the augmented feature point or augmented point. In some embodiments of the present invention, the augmented feature point may be the intersection of two lines defined by the feature points that reinforces the center region of the ID document. In some embodiments of the present invention, feature points may be augmented to cover regions of the ID document where there is a lack or absence of existing landmarks. An advantage of the embodiments of the present invention is that the augmented feature points significantly enhance and improve the accuracy of perspective transformation by increasing the number of reliable feature points and the coverage of the targets on the ID document. Thereby, another advantage of the embodiments of the present invention is that superior dewarped results can be expected on account of introduction of more feature points located at different parts of

the target. Embodiments of the present invention may involve any method for augmenting the feature points. In some embodiments of the present invention, a simple linear algebra concept may be applied for augmenting feature points. In some embodiments of the present invention, advanced methods such as affine transformation may be applied in feature point augmentation.

[0091] Thereafter, as at block **1060**, the ID document is dewarped using perspective transformation of the extracted feature points, that is, feature points and the augmented feature points. In some embodiments of the present invention, when the feature points extracted at block **1040** sufficiently cover all the areas of the ID document, perspective transformation may be performed using only said feature points. In some embodiments of the present invention, when the feature points extracted at block **1040** partially cover the areas of the ID document, perspective transformation may be performed using the feature points and augmented feature points extracted at block **1050**. In some embodiments of the present invention, dewarping is performed using the traditional image processing unit, while the process landmark segmentation, shape approximation, feature point extraction and feature point augmentation, as disclosed at blocks **1010, 1020 , 1030 , 1040** and **1050** are obtained using the deep learning neural networks or techniques.

[0092] **FIG. 11** is a flow diagram that illustrates an example method **1100** for segmenting landmarks in an ID document, in accordance with the disclosed embodiments. As disclosed in **FIG. 11**, the method **1100** may be implemented by extracting, at least one image feature in each landmark in the ID document, as at block **1110**. In some embodiments of the present invention, the extraction may be conducted by an encoder of a neural network model, for example, U-net model. Then as at block **1120**, localization of the image feature may be performed. In some embodiments of the present invention, the localization may be conducted by a decoder of the neural network model. Thereafter, as disclosed at block **1130**, the method **1100** may be implemented by capturing and labeling the image feature as the segmented landmark.

[0093] **FIG. 12** is a flow diagram that illustrates an example method **1200** for extracting feature points of a segmented landmark with a shape similar to a rectangular or a parallelogram on an ID document, in accordance with the disclosed

embodiments. As disclosed in **FIG. 12**, the method **1200** may be implemented by approximating the segmented landmark to a rectangular shape, as at block **1210**. In some embodiments of the present invention, the segmented landmark is approximated to a rectangle using a process of a rectangular shape approximation by the neural network model, for example U-net model. Then, as at block **1220**, the shape approximated landmark is obtained. Then as at block **1230**, a midpoint on each edge of the shape approximated landmark is calculated. Then as at block **1240**, an axis along two opposite midpoints of the edges are formed. Thereafter, as at block **1250**, a centroid is obtained at an intersection feature point of the axes. Then as at block **1260**, the feature points comprising the midpoints and the centroid are extracted.

[0094] **FIG. 13** is a flow diagram that illustrates an example method **1300** for extracting feature points of a segmented landmark with a shape similar to an ellipse on an ID document, in accordance with the disclosed embodiments. As disclosed in **FIG. 13**, the method **1300** may be implemented by approximating the segmented landmark to an ellipse or a circle, as at block **1310**. In some embodiments of the present invention, the segmented landmark is approximated to a circular or an elliptical shape using a process of an elliptical shape approximation by the neural network model, for example U-net model. Then, as at block **1320**, the shape approximated landmark is obtained. Then as at block **1330**, major and minor axes of the ellipse are formed. Thereafter, the method **1300** may be implemented by obtaining the midpoints on the ellipse using intersection of the major and minor axes, as at block **1340**. Further, as at block **1350**, a centroid of the ellipse is obtained using midpoints on the ellipse. As at block **1360**, the feature points comprising the midpoints and the centroid are extracted.

[0095] **FIG. 14** is a flow diagram that illustrates an example method **1400** for obtaining augmented feature points from the feature points that are extracted from the methods **1200** and **1300** of **FIG. 12** and **FIG. 13** respectively, in accordance with the disclosed embodiments. In one embodiment of the invention, the augmented feature point is obtained by calculating one or more interpolated points of the feature points in an upper and/or lower region of the landmarks in the identity document, using a linear interpolation. As disclosed in **FIG. 14**, the method **1400** may be implemented by

calculating a centroid of the feature points of the segmented landmarks, as at step **1410**. Then, the method **1400** may be implemented by calculating an intersection point of lines defined by the feature points of the segmented landmarks. In one embodiment of the invention, the augmented feature point comprises said one or more interpolated points, the centroid and the intersection point.

[0096] **FIG. 15** is a flow diagram that illustrates an example method **1500** for obtaining a dewarped image of an ID document using perspective transformation, in accordance with the disclosed embodiments. As at block **1510**, a reference image of the ID document with a plurality of reference points corresponding to the augmented feature points are prepared. For each type of ID document, a reference image *R* with a plurality of reference points corresponding to the feature points of the ID document to be aligned is prepared. Then as at block **1520**, a projection matrix is computed using the plurality of reference points and a set of augmented feature points. In embodiments of the present invention, the projection matrix *H* is computed from formula $P' = HP$, where *P* is equal to $\{p_1, p_2, \ldots, p_k\} + \{g_1, g_2, \ldots, g_l\} + \{m_1, m_2, \ldots, m_n\}$. *P* is the extracted feature point combination including the feature points from the landmarks and augmented points. *P'* is equal to $\{p_1', p_2', \ldots, p_k'\} + \{g_1', g_2', \ldots, g_l'\} + \{m_1', m_2', \ldots, m_n'\}$ and refers to the corresponding feature points in reference image *R* displayed.

[0097] Thereafter, as at block **1530**, the dewarped image of the ID document is obtained by multiplying the projection matrix with the ID document. The final dewarped image *I'* can be obtained by multiplying the projection matrix *H* by the image *I*. The equation for obtaining the final dewarped image *I' (I'= HI)* is equal to the product of the projection matrix *H* and the image *I*. The system crops out the region where the ID is mapped to in *I'*. The cropped region is the final dewarped image *I'* having the same size as the reference image *R.*

[0098] As disclosed above, the reference image is obtained from a real flatten ID with standard layout. The location of the reference points are collected as the target points P'. For any new image, the feature points at corresponding locations are collected using a neural network model, for example a U-Net model, and mathematical augmentation is performed. The locations of the feature points are denoted set *P*. The projection matrix *H* shows the way to convert *P* to *P'*. Further, the reference image *R*

provides the information pertaining to the placement and positioning of the significant points in a successfully aligned ID document. An advantage of the present disclosure is that the extraction of feature points of each landmark is conducted along with extraction of corresponding augmented feature points, thereby eliminating discrepancies in alignment on account of miscellaneous and external factors such as pixel information, like the color or texture of the ID document. Accordingly, embodiments of the present invention provide a robust solution to document alignment of captured ID documents having different brightness, orientations and/ or resolutions, or being of a lower quality and suffering from blurriness, reflection and/ or insufficient brightness. Additionally, embodiments of the present invention eliminate any dependence of the extraction of feature points in the reference images on other variant factors such as layout of the ID documents, and therefore, embodiments of the present invention are able to efficiently and aptly align ID documents, thereby expanding the scope of application of the disclosed solution. For the above reasons, the alignment solution of the present disclosure is capable of being adapted for any recognition tasks such as facial recognition requiring image alignment for subsequent processing. In some embodiments of the present invention, the disclosure can be adapted to segment the landmarks such as eyes, noses, and mouth and augment the feature points to cover the entire face. The frontal face can be obtained by the perspective transformation based on the feature points.

[0099] Any combination of feature points can be the reference sets for alignment as long as the way to extract the features points are the same in the new image and the reference image. For example, if p_1 in the new images is chosen as one of the feature points, then the corresponding p_1' in reference image should also be chosen. The feature points in the new image and reference image are mapped one to one.

[00100] The present invention also discloses a computer readable medium storing a program comprising instructions that, when executed by a system, cause the system to perform the method for aligning at least one ID document. In an embodiment of the invention, one or more computing devices comprising the computer readable medium is used to perform the methods **1000**, **1100**, **1200**, **1300**, **1400** disclosed in **FIG. 10**, **FIG. 11**, **FIG. 12**, **FIG. 13, FIG. 14** and **FIG. 15**, respectively.

[00101] The present invention has many advantages. In the feature point extraction stage, the system applies deep-learning based semantic segmentation trained with enormous data samples of different brightness, orientations and resolutions to extract the feature points of the input image. It is more robust than traditional image processing techniques such as SIFT and SURF which relies on excellent image quality to achieve acceptable performance.

[00102] Matching the feature points in the input image with those in the reference image is difficult when using classical matching scheme in the image processing. Image processing techniques, for example Random Sample Consensus (RANSAC), apply global mapping, so that the points sharing similar features at different locations could be wrongly matched. In contrast, the present invention captures the feature points from each landmark and augmented based on the selected known feature points. The feature points are extracted in a certain order that make the matching straightforward without any calculation involved.

[00103] As the order of the extracted feature points based on traditional algorithms such as SIFT, SURF and rotated robust independent elementary features (ORB) is inconsistent, consequently, there is no straightforward way to augment the feature points for regions where there is a lack of significant landmarks to provide feature points while it is fairly trivial to append new feature points by simple mathematical calculation in the proposed approach.

[00104] The semantic segmentation or the neural network model in the present disclosure is a relatively small neural network that can be properly trained with a smaller training set. In addition, the perspective transformation is done by the traditional image processing method which consumes less computing power. It is more efficient in terms of speed compared with the pure deep learning solutions.

[00105] The proposed system applies a generic pipeline that can be applied to not only various types of ID cards but also other recognition tasks such as facial recognition that requires image alignment for subsequent processing. The system can be modified to segment the landmarks such as eyes, noses, and mouths and augment

the feature points to cover the entire face. The frontal face can be obtained by the perspective transformation based on the feature points.

[00106] It will be appreciated that variations of the above disclosed and other features and functions, or alternatives thereof, may be desirably combined into many other different systems or applications. Also, various presently unforeseen or unanticipated alternatives, modifications, variations or improvements therein may be subsequently made by those skilled in the art which are also intended to be encompassed by the following claims.

[00107] Although embodiments of the current disclosure have been described comprehensively in considerable detail to cover the possible aspects, those skilled in the art would recognize that other versions of the disclosure are also possible.

CLAIMS

What is claimed is:

1. A system for aligning at least one identity document comprising:

a neural network model for obtaining one or more segmented landmarks in the identity document,

wherein the identity document is captured at one or more angles, at one or more locations such that said one or more segmented landmarks in the identity document are obtained by performing a landmark segmentation process; and

an image processing unit for processing the identity document by performing shape approximation on said one or more segmented landmarks, extracting a plurality of feature points of said one or more segmented landmarks and dewarping the identity document by performing a perspective transformation of said plurality of feature points.

2. The system of claim 1, wherein the image processing unit augments said plurality of feature points to cover at least one non-landmark area to obtain at least one augmented feature point and dewarping the identity document by performing a perspective transformation of said plurality of feature points and said at least one augmented feature point.

3. The system of claim 2, wherein in the perspective transformation, a projection matrix is computed using a plurality of reference points of said plurality of feature points and said at least one augmented feature point on a reference identity image and the identity document is dewarped by multiplying the projection matrix with the identity document.

4. The system of claim 1, wherein the neural network model is a U-net model.

5. The system of claim 1, wherein the neural network model is trained using a plurality of image pairs, each pair consisting of an original image and a corresponding labeled image captured at one or more angles and at one or more locations.

6. The system of claim 1, wherein the neural network model comprises at least

one encoder for extracting at least one image feature in said one or more segmented landmarks in the identity document and at least one decoder for performing localization of said at least one image feature.

7. The system of claim 1, wherein said plurality of feature points are extracted from a shape approximated landmark.

8. The system of claim 2, wherein said at least one augmented feature point is obtained by calculating one or more interpolated points of said plurality of feature points in an upper and/ or lower region of the landmarks in the identity document, using a linear interpolation.

9. The system of claim 8, wherein said at least one augmented feature point is obtained by calculating:

at least one centroid of said plurality of feature points of said one or more segmented landmarks; or

an intersection point of the lines defined by said plurality of feature points of said one or more segmented landmarks.

10. The system of claim 9, wherein said at least one augmented feature point comprises one or more interpolated points, the centroid and the intersection point.

11. A method for aligning at least one identity document comprising,

obtaining, by a neural network model, at least one landmark of the identity document at one or more angles, at one or more locations;

performing, by the neural network model, at least one segmentation of said at least one landmark in the identity document to obtain one or more segmented landmarks;

performing, by the image processing unit, at least one shape approximation on said one or more segmented landmarks to obtain at least one shape approximated landmark;

extracting, by the image processing unit, a plurality of feature points from said one or more segmented landmarks;

and

dewarping, by the image processing unit, the identity document using perspective transformation of said plurality of feature.

12. The method of claim 11, further comprises:

extracting, by the image processing unit, at least one augmented feature point from said plurality of feature points to cover at least one non-landmark area; and

dewarping, by the image processing unit, the identity document using perspective transformation of said plurality of feature points and said at least one augmented feature point.

13. The method of claim 12, wherein said one or more segmented landmarks is obtained by:

extracting at least one image feature in said at least one landmark in the identity document using an encoder;

performing localization of said at least one image feature using a decoder; and

capturing and labeling at least one image feature as said one or more segmented landmarks.

14. The method of claim 13, wherein said plurality of feature points are obtained by:

performing, by the image processing unit, at least one rectangular approximation on said one or more segmented landmarks;

obtaining said at least one shape approximated landmark;

calculating a midpoint on each edge of said at least one shape approximated landmark;

forming an axis along two opposite midpoints of the edges;

obtaining a centroid at an intersection feature point of the axis; and

obtaining said plurality of feature points of the one or more segmented landmarks comprising the midpoints and the centroid.

15. The method of claim 13, wherein said plurality of feature points are obtained by:

performing, by the image processing unit, at least one elliptical approximation on said one or more segmented landmarks;

obtaining said at least one shape approximated landmark;

forming a major and a minor axis of the ellipse;

obtaining the midpoint on the ellipse using intersection of the major and minor axes and a contour of the ellipse;

obtaining the centroid of the ellipse using the midpoints on the ellipse; and

obtaining said plurality of feature points of the one or more segmented landmarks comprising the midpoints and the centroid.

16. The method of any one of claim 14 or 15, wherein said at least one augmented feature point is obtained by calculating one or more interpolated points of said plurality of feature points in an upper and/or lower region of the landmarks in the identity document, using a linear interpolation.

17. The method of any one of claims 14 to 16, wherein said at least one augmented feature point is obtained by:

calculating at least one centroid of said plurality of feature points of said one or more segmented landmarks; and

calculating an intersection point of lines defined by said plurality of feature points of said one or more segmented landmarks.

18. The method of any one of claims 14 to 17, wherein said at least one augmented feature point comprises said one or more interpolated points, the centroid and the intersection point.

19. The method of any of claims 11 to 18, wherein dewarping the identity document using perspective transformation of said plurality of feature points and said at least one augmented feature point comprises the steps of:

obtaining a plurality of reference points of a reference image of the identity document, corresponding to said plurality of feature points and said at least one augmented feature point;

computing a projection matrix using the series of said plurality of reference points and said plurality of feature points and said at least one augmented feature

point; and

obtaining a dewarped image of the identity document by multiplying the projection matrix with the identity document.

20. A computer readable medium storing a program comprising instructions that, when executed by a system, cause the system to perform the method for aligning at least one Identity (ID) document, of any of claims 11 – 19.