Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 41 (2008) 1867–1877
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

Identifying elephant photos by multi-curve matching

A. Ardovini, L. Cinque, E. Sangineto*

Computer Science Department, University of Rome "La Sapienza", via Salaria 113, 00198 Rome, Italy

Received 10 January 2007; received in revised form 15 October 2007; accepted 12 November 2007

Abstract

We present in this paper an elephant photo identification system based on the shape comparison of the nicks characterizing the elephants' ears. The method we propose can deal with very cluttered and noisy images as the ones commonly used by zoologists for wild elephant photo identification. Difficult segmentation problems are solved using rough position information input by the user. Such information is used by the system as a basis for a set of segmentation and normalization hypotheses aiming at comparing a query photo Q with different photos of the system's database possibly representing the same individual as Q. The proposed shape comparison method, based on matching multiple, non-connected curves, can be applied to different retrieval by shape problems. Examples with real wild elephant photos are shown.

Keywords: Elephant photo identification; Image retrieval by shape; Symmetric invariant multiple curve matching

1. Introduction

This work has been motivated by the necessity to automate as much as possible the currently tedious and time-consuming human-made photo recognition of individuals in an elephant population. In order to study demographic processes of the elephant populations, as well as to plan their conservation, it is essential to trace the elephant movements along time. Photographic elephant identification is one of the cheapest, less invasive and most common marking system. It is currently a human-made operation based on the comparison of each new photo with all the images of the zoologist's database (usually composed of hundreds of exemplars) looking to possible similarities with respect to some distinguishing biological features. One of these features is the shape of the nicks which characterize the borders of the ears (e.g., see Fig. 1). Such nicks can be used to uniquely identify an exemplar almost like fingerprints for human beings.

For a computer vision point of view the problem is not trivial because of the commonly highly cluttered and low-contrast images representing the interesting shapes. In fact, photos of wild elephants can contain trees and shrubs either in the background or partially occluding the animal's ear, while the thick veins of the elephant's skin as well as the low contrast between the ear surface and the body skin make the segmentation of the ear's border from the surrounding lines a very hard process (e.g., see Fig. 3). We have dealt with this problem by building a semi-automatic system in which the human operator approximately selects the nick position using the mouse and the system produces a set of hypotheses on the nick composition based on different candidate endpoints. These hypotheses help to refine the segmentation–detection process and are verified by taking into account both their positions with respect to the whole ear and the nick shape by matching the selected lines with the nick curves contained in the system's database. The output of the system is the ranked list of the top 10 most similar individuals to the input query.

The proposed system is currently used by some researchers of the Zoological Department of the University of Rome "La Sapienza", who are studying the African elephant population of Zakouma Ciad National Park [1]. The system has also helped the zoologists in discovering a few photos of the same individuals miss-classified as different exemplars during the previous hand-made identification attempts. The experimental results presented in Section 7, as well as all the example photos shown in this article have been taken from their databases.

* Corresponding author. Tel.: +39 06 49918358; fax: +39 06 8541842.
E-mail address: sangineto@di.uniroma1.it (E. Sangineto).

0031-3203/$30.00
doi:10.1016/j.patcog.2007.11.010

Fig. 1. Some examples of elephant photos. Top row: the nicks are marked in yellow. Sometimes elephant identification using the nick shape is a non-trivial task also for human beings (e.g., bottom row). Every photo here represents a different individual.

The rest of the paper is organized as follows. In Section 2 we briefly review some very recent proposals for animal photo identification and other general approaches to curve matching. In Section 3 we show the preprocessing and symmetric invariant normalization operations performed by the system on each image. The matching algorithm is presented in Sections 4 (position matching) and 5 (shape matching). In Section 6 we discuss the computational costs of our proposed approach and we compare its recognition peculiarities with other methods. Finally, we show our experimental results in Section 7 and we conclude in Section 8.

2. Related works

In this century different computer vision techniques have been proposed for semi-automatic photo identification of animals, with a particular attention to marine mammals. To the best of our knowledge all the systems so far proposed are based on two main principles. First, the appearance of (a part of) the analyzed animal is semi-automatically segmented in order to isolate, with the help of a human operator, those distinctive biological feature(s) which make it possible the animal identification. The intervention of a human operator in this phase is necessary due to the usually cluttered, low-contrast and very noisy images of wild animals in uncontrolled environments which make completely automatic feature detection a non-reliable task. Second, once the feature(s) have been isolated, they must be suitably represented and compared with the previously stored feature representations of the system's database animals. The feature representation should be invariant to rigid geometric transformations (e.g., translations, rotations, scale changes, etc.) in order to deal with changes in the camera point of view.

In Refs. [2–4] the authors deal with photo identification of, respectively, humpback whales and gray whales. The features they use are the white/light gray patches of the whales' tails. In Refs. [2,3] an interactive segmentation process is used to segment the fluke region in black and white pixels. Three fixed landmarks (the left and right fluke tips and the middle notch) are used to compute a reference system invariant with respect to rotations, translations and scale changes. Using these three points a grid is superimposed on the whole fluke region which is then partitioned with respect to the grid's elements. An n dimensional feature vector is extracted to represent the fluke. The ith element of such a vector is given by the ratio of white pixels to the total number of pixels of a corresponding element of the grid. Patterns extracted from this feature vector are used for fluke matching. In Ref. [4] affine transformations of the fluke with respect to the camera position are dealt with by using Zernike moment invariants to represent the fluke's white patches.

Dolphin identification is achieved in Ref. [5] using the characteristic shape of the dorsal fin. The curve representing a profile view of a dorsal fin is extracted from the image using an interactive edge detection tool. The notches characterizing the fin are used in order to identify the individual. Therefore, the biological features used are similar to the elephant case. The authors propose to represent the extracted curve using an attributed string, in which each character (*primitive*) represents a curvature element and is associated with some attributes (e.g., length, enclosed area, etc.). A string matching procedure is used in order to estimate the similarity between two curves. However, it should be noted that the extraction of symbolic primitives from a digital shape is sensitive to noise and quite unstable and brings to a possible information loss [6].

A similar problem is dealt with in Ref. [7]. The characteristic curves of dolphin fins, sea lion flippers and gray whale flukes are used to, respectively, recognize individuals of the three cetacean species. In this case all the points of a given curve are used in the matching process without information loss. The query curve (Q) is compared with each curve (C)

of the system's database. Given a pair of curves Q and C, the points corresponding to the respective curvature maxima and minima are used as landmarks in order to compute the affine transformation which best aligns Q upon C. The "mismatch area" between the two aligned curves is then computed and used to estimate the curve dissimilarity. Unfortunately, there are some assumptions on the number and on the nature of the curvature maxima and minima (the landmarks used to compute the alignment) which makes the approach dependent on the particular animal anatomy. Moreover, both Q and C need to be unoccluded and uninterrupted. For instance, if Q and C were piecewise connected lines composed of connected subcurves $Q_1, Q_2, \ldots, Q_N$ and $C_1, C_2, \ldots, C_M$, it would not be clear how to match the single components and to compute the overall mismatch area.

A similar problem is present in most of the approaches so far proposed for animal photo identification by shape matching. For instance, the extraction of a connected curve is also necessary in Ref. [8] to classify Sperm Whale flukes. Once isolated from the background, the trailing edge of the animal's tail is a curve dealt with as a mono-dimensional signal. The authors apply a wavelet transform to such signals and compare the resulting coefficients.

Unfortunately, the border of an elephant's ear profile extracted using a common edge detector is very hardly completely connected (this situation never happened in our experiments) and it is sometimes partially occluded by other elephants or trees (e.g., Figs. 1–3). For this reason we need to deal with partially defined curves. Moreover, it is not the shape of the whole ear which is important for the recognition task, but the shape details of each single nick, as well as the nick position with respect to the rest of the ear. Generally speaking, the recognition process must be based on the comparison of two pairs of non-connected curve sets $\{Q_1, Q_2, \ldots, Q_N\}$ and $\{C_1, C_2, \ldots, C_M\}$.

In computer vision and pattern recognition literature curve matching problems are usually dealt with using the assumption of a closed or at least connected, single curve. *Curvature scale space* [9–11] approaches are based on a scale-depending representation of a given curve C obtained by progressively smoothing C. C is assumed to be connected. *Fourier descriptors* or *wavelet coefficients* are often used to describe the external border of a connected silhouette [12,8]. *Shape signature*-based approaches represent a curve using a string (called *signature*), in which each element (*character* or *primitive*) represents an atomic piece of the line (e.g., see Ref. [5]). Also shape signatures cannot deal with partially defined curves.

An *active contour model* (also called *snake*) [13,14] is a deformable curve (S) which can be superimposed upon an image edge map I and progressively deformed until either the points of S and I overlap on each other or the deformation of S exceeds a given value. Snakes and their statistical learning-based version *point distribution models* [15] can in principle deal with partial occlusions and/or missing points in I. Nevertheless, S needs to be connected which introduces problems in its extraction from a cluttered image. Moreover, all the points of S have the same importance in the matching phase. Last but not the least, a cluttered background in the proximity of the elephant's ear (e.g., see Fig. 3) can easily trap the deformation process into a false local minimum.

Contrary to the above mentioned methods, the generalized Hough transform (GHT) [16] is a well-known technique with a proved ability in dealing with non-connected curves. A non-analytic curve C is represented by means of an "R-Table" T which contains the translation offsets of each point of C with respect to a given reference point $\mathbf{p}$. At searching time, C is looked for in the image edge map I by computing the possible positions of $\mathbf{p}$ with respect to all the points of I using T. No continuity assumption is used in this case, and in fact the GHT and its numerous, different versions (e.g., see Ref. [17]) are used in problems in which occlusions and partial data prevent an exact segmentation of the image edge map. Nevertheless, there are two main reasons for which the GHT cannot be used in our domain. First, we need to take into account not only x and y translations of the query curves with respect to the database curves, but also rotational and scale parameters, due to possible viewpoint changes (e.g., see Fig. 1). It is well known that the GHT efficiency is quite critical in dealing with such additional parameters. Moreover, the typical image background of an elephant in its natural habitat is composed of very thick twigs and shrubs (see Fig. 3) which make very confusing the voting phase. We have implemented the traditional version of the GHT with only translation parameters and applied it to a subset of fixed scale and orientation images of elephants. Our experimental trials have shown that, as we expected, false positives are very frequently detected on the elephant's skin vein or on the image background regions.

The method proposed in this paper is based on two main steps. We assume to compare a set $Q_1, \ldots, Q_N$ of curves representing the visible nicks of the query image and a set $P_1, \ldots, P_M$ of curves of a given individual of the database. First, we globally compare $Q_1, \ldots, Q_N$ and $P_1, \ldots, P_M$ searching for a matching of a subset of curves having consistent positions with respect to a few reference points on the animal's head selected by the user (Section 4). Second, each valid global match is subsequently locally verified by separately comparing pairs of matched curves (Section 5). In the global step the ear's border between Q_i and Q_{i+1}, as well as the border between P_j and P_{j+1} do not need to be defined. This makes the comparison process depending only on the nicks. Conversely, in locally comparing the shape of two given nicks Q_i and P_j, both Q_i and P_j must be connected. The connectivity property of the curve shape of each single nick can be easily guaranteed by a minor human intervention during the interactive segmentation process as described in the next section.

3. Extraction of the nick curvature

In this section we show the different phases of the semi-automatic segmentation process performed on each image when it is either off-line stored in the system's repository or used as on-line query.

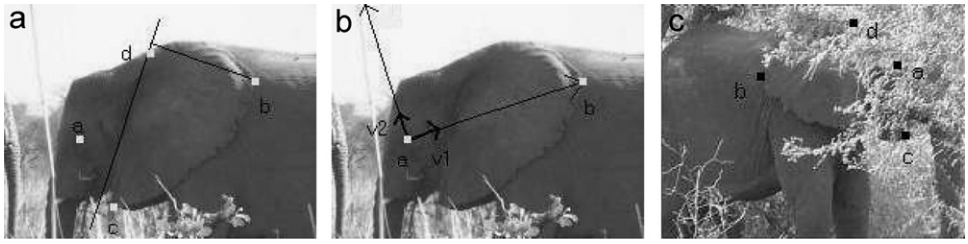

Fig. 2. The ear reference system. (a) Points **a**–**d** selected by the user. (b) The symmetric-invariant reference system built using points **a** and **b**. (c) Points **a**–**d** approximately selected on a difficult image with a large occlusion.

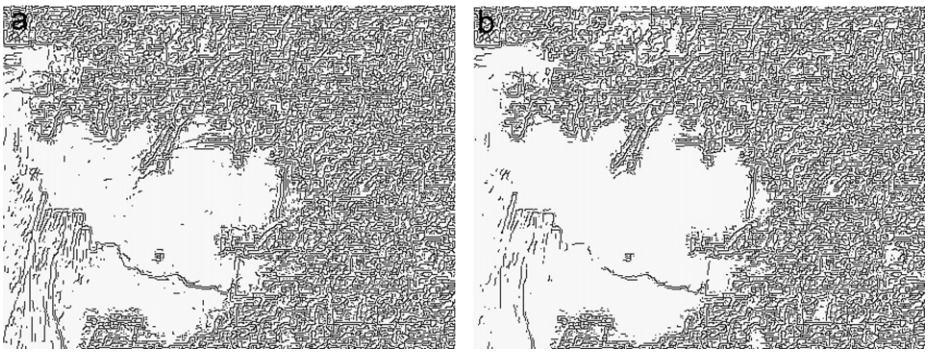

Fig. 3. The edge detection of Fig. 2(c) with slightly different binarization threshold values selected by the user. Note that in (b) a lower threshold value makes the nick on the ear's border no more visible although the very cluttered background still remains.

First of all, the user is requested to input a few reference points in order to approximately define the elephant's head position, orientation and scale. With the assumption that the ground direction is given by the bottom of the image, only two points are sufficient to define a symmetric-invariant reference system for the whole head and to automatically select the right/left profile. We chose to represent the nicks in a symmetric-invariant reference framework rather than in an affine-invariant one for two reasons. First, the images used by zoologists for elephant photo identification are usually profile photos, with the ears very close to the animal's body, thus out-of-plan ear rotations are very uncommon. Second, generic affine transformations would have required the user selection of three rather than two points, making the system's precision more dependent on the accuracy used by the human operator in choosing the reference points. As we will see below, the definition of fixed and visible reference points on a generic elephant ear is not a trivial task.

In fact the user is asked for clicking on the eye's central position (**a**) and on point **b** which is defined as follows. Imagine to trace the line delimiting the attachment of the ear with the head (see Fig. 2(a)). **b** is the point with the maximum distance from this line. When either **a** or **b** are not visible due to occlusions or other reasons, the user can click on an approximate estimation of their positions (e.g., Fig. 2(c)). Since the definition of **b** is a bit arbitrary and since different users can select different points **a** and **b** for the same image, the spatial information so obtained can only be approximately used in order to estimate the nicks' positions with respect to the whole ear. For redundancy reasons, the system allows the user to select other two points: **c**, defined as the lowest point of the lobe, and **d**, the highest attachment point of the ear with the head (Fig. 2(a)). **c** and **d** are stored together with the image in the system's repository. At running time, the user can choose to input a query image selecting **a** and one among **b**–**d**. However, our experimental trials have proved that **b** gives the most stable results and the reason is that **c** is often non-visible due to frequent lobe folds while the position of **d** is sometimes unclear because of the low-contrast between the ear and the body. Without loss of generality in the following we will refer to only **a** and **b**.

Once **a** and **b** have been selected, the system performs an edge detection operation using the standard Canny edge detector [18] whose binarization threshold can possibly be varied by the user by means of a suitable slider bar provided by the system's interface (e.g., compare Figs. 3(a) and (b)). Let us call $\overline{I}$ the edge map of I. After that, the user is requested to input the initial (**e**) and the final (**f**) endpoint for each nick (see Fig. 6). The order in which **e** and **f** are inserted is not important. Once a pair of endpoints for every nick has been input, the system automatically computes the shortest path in $\overline{I}$ between **e** and **f**. If **e** and **f** are not connected or the computed path is not the desired one, the user can add/delete points on $\overline{I}$ using a suitable GUI. This is the only time-consuming human intervention requested by our system, which is only occasionally necessary for few points per nick. However, this manual operation takes few dozens of seconds per nick.

The shortest path in $\overline{I}$ between **e** and **f** gives the set of points defining the current nick N (Fig. 6). Suppose: $N = \{\mathbf{p}_1, \ldots, \mathbf{p}_r\}$ (where $\mathbf{p}_1 = \mathbf{e}$ and $\mathbf{p}_r = \mathbf{f}$). N is subsampled using a fixed number

of points n (currently, $n = 25$) using the following algorithm:

Subsampling Procedure ($\{\mathbf{p}_1, \ldots, \mathbf{p}_r\}$)
1 $step := \frac{r-1}{n-1}$, $s := step + 1$.
2 $P[1] := \mathbf{p}_1$, $P[n] := \mathbf{p}_r$.
3 For $i \in [2, n-1]$, do:
4 $k := \lfloor s \rfloor$, $d := s - k$.
5 $P[i] := InterpolatedPoint(\mathbf{p}_k, \mathbf{p}_{k+1}, d)$.
6 $s := s + step$.
7 Return P.

In Step 4 "$\lfloor s \rfloor$" is the floor operation on s and, hence, $0 \leqslant d < 1$ (i.e., k is the greatest integer smaller than or equal to s and d is the remaining decimal fraction). *InterpolatedPoint*() in Step 5 is defined as follows:

$InterpolatedPoint(\mathbf{p}_k, \mathbf{p}_{k+1}, d) = \mathbf{p}_k + d(\mathbf{p}_{k+1} - \mathbf{p}_k)$.

Once all the nicks have been acquired by the system, the current image can either be stored in the system's database or used as a query in order to retrieve possible different photos of the same individual previously stored in the database. In the first case the original gray-level image I is stored in the repository together with the points $\mathbf{a}_I$ and $\mathbf{b}_I$ and the set of nicks $P_1, \ldots, P_M$ computed as described above. Each P_i is in turn associated with the user-selected endpoints $\mathbf{e}_i$ and $\mathbf{f}_i$, which are normalized with respect to a symmetric-invariant reference system defined using the reference points $\mathbf{a}_I$ and $\mathbf{b}_I$. The normalization is performed as follows. Assume $\mathbf{x} = \mathbf{b}_I - \mathbf{a}_I$, and $d = \|\mathbf{x}\|_2$. Moreover, let $\mathbf{v}_1 = \frac{1}{d}\mathbf{x}$; and $\mathbf{v}_2$ be an unitary vector centered in $\mathbf{a}_I$ and orthogonal to $\mathbf{v}_1$ whose direction is fixed swapping a $\pi/2$ angle clockwise from $\mathbf{x}$ for the right ear and counterclockwise for the left ear (see Fig. 2(b)). If $\mathbf{e}$ is an endpoint selected by the user, the normalized nick endpoint $\mathbf{e}'$ is then given by

$$\mathbf{e}' = \frac{1}{d}\begin{pmatrix} \mathbf{v}_1^{\mathrm{T}}(\mathbf{e} - \mathbf{a}_I) \\ \mathbf{v}_2^{\mathrm{T}}(\mathbf{e} - \mathbf{a}_I) \end{pmatrix} \tag{1}$$

and the same is for $\mathbf{f}'$. From now on we will refer to $\mathbf{e}'$, $\mathbf{f}'$ and $\mathbf{e}$, $\mathbf{f}$ as, respectively, the normalized and the original endpoint representations.

The reference system defined by $\mathbf{v}_1$ and $\mathbf{v}_2$, centered in $\mathbf{a}_I$ and with unit of length d is invariant with respect to symmetric transformations of the head appearance. In fact the same head can be represented in different photos with different translation offsets, scale factors and in-plane rotations (i.e., tilt movements of the head with respect to the body, like in Fig. 1, bottom row, leftmost image). Note that three points on the ear (e.g., $\mathbf{b}_I$, $\mathbf{c}_I$ and $\mathbf{d}_I$) would be enough to deal with out-of-plan rotations using an affine-invariant reference system [19,7]. Nevertheless, for the reasons above mentioned we prefer to use only two reference points and to deal with only symmetric transformations which are enough for the type of photos used in the elephant's identification domain.

The same class of rigid symmetric transformations is also taken into account when pairs of curves representing single nicks are compared (Section 5). Also in this case, two reference points are enough for the type of analyzed images, and they can be naturally chosen to be the nick's endpoints. As we will see in Section 5, each nick of I is separately compared with the nicks of Q. To this aim the points of each single nick P_i are normalized with respect to the P_i's endpoints. Assume that $\mathbf{e}$ is the higher and $\mathbf{f}$ the lower endpoint of the nick $P = \{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$, being $\mathbf{p}_1 = \mathbf{e}$ and $\mathbf{p}_n = \mathbf{f}$. Then we can define: $\delta = \|\mathbf{f} - \mathbf{e}\|_2$, where $\|\mathbf{x} - \mathbf{y}\|_2$ is the Euclidean distance between the points $\mathbf{x}$ and $\mathbf{y}$; $\mathbf{w}_1 = 1/\delta(\mathbf{f} - \mathbf{e})$ and $\mathbf{w}_2$ be an unitary vector centered in $\mathbf{e}$ and orthogonal to $\mathbf{w}_1$ whose direction is fixed swapping a $\pi/2$ angle clockwise from $\mathbf{w}_1$ for the right ear and counterclockwise for the left ear. The normalized nick $P' = \{\mathbf{p}'_1, \ldots, \mathbf{p}'_n\}$ is obtained by imposing, for each $\mathbf{p}'_j$ $(1 \leqslant j \leqslant n)$ that:

$$\mathbf{p}'_j = \frac{1}{\delta}\begin{pmatrix} \mathbf{w}_1^{\mathrm{T}}(\mathbf{p}_j - \mathbf{e}) \\ \mathbf{w}_2^{\mathrm{T}}(\mathbf{p}_j - \mathbf{e}) \end{pmatrix}. \tag{2}$$

Finally, if the image is used as query (Q), we indicate with $Q_1, \ldots, Q_N$ its nicks, being each Q_i $(1 \leqslant i \leqslant N)$ associated with the normalized endpoints $\mathbf{g}'_i$ and $\mathbf{h}'_i$, obtained using Eq. (1) and the corresponding reference points $\mathbf{a}_Q$ and $\mathbf{b}_Q$. We will see in Section 5 how also the inner points of each nick Q_i are normalized using the same procedure described for P_i, which will make it possible a direct point-to-point comparison between nicks of different images.

4. Curve sequence matching

Given a database image I with M nicks $(P_1, \ldots, P_M)$ and a query image Q with N nicks $(Q_1, \ldots, Q_N)$, the problem is to find a portion of I and a portion of Q which are similar, i.e., contain the same nicks. Even if both I and Q represent the same elephant's ear E, we can have $N \neq M$. In fact, a portion of E can be partially occluded either in I or in Q. Moreover the lobe and other parts of the ear are sometimes folded and non-visible.

We assume in the following that, if both Q and I represent the same ear E of the same elephant, then there is only one portion V of E which is visible in both I and Q (e.g., see Figs. 1 and 2(c)). V is defined by a subsequence of nicks

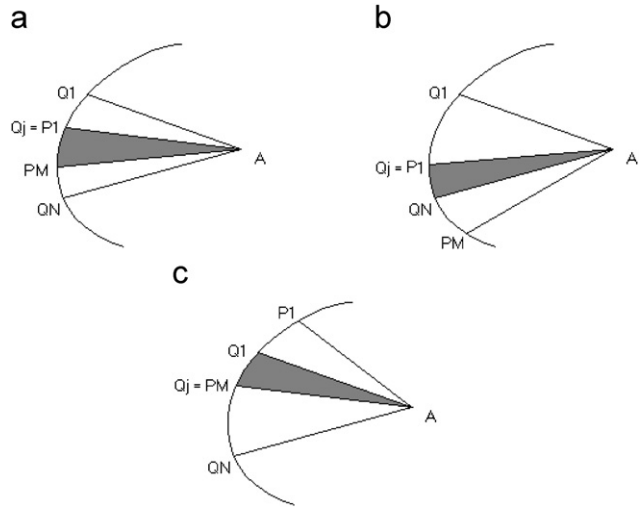

Fig. 4. Position matching of two nick sequences. The visible common subsequence V is marked in gray. (a) $(P_1, \ldots, P_M)$ is completely included in $(Q_1, \ldots, Q_N)$. (b) V is delimited by P_1 and Q_N. (c) V is delimited by Q_1 and P_M.

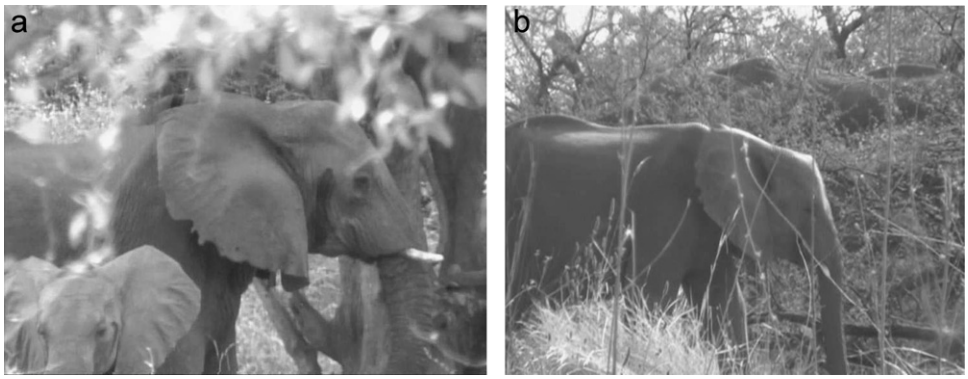

Fig. 5. The first two individuals ranked by the system with respect to the query of Fig. 2(c). The first image corresponds to the same individual as the query.

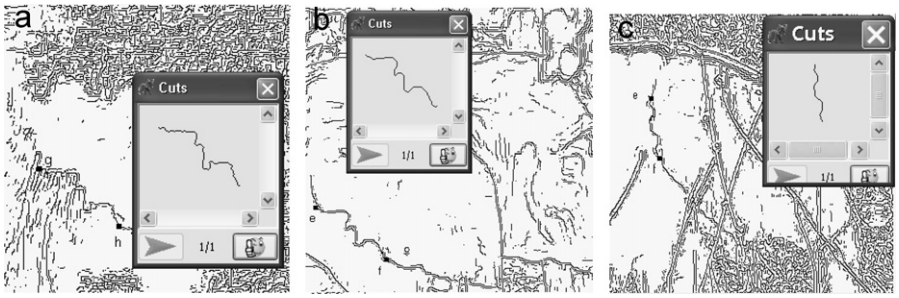

Fig. 6. Nick curve extraction. (a) The nick extracted from the image of Fig. 3(a), used as query. (b) and (c) The nicks extracted, respectively, from the images of Figs. 5(a) and (b) when these were stored in the system's database.

common to both the sequences $(P_1, \ldots, P_M)$ and $(Q_1, \ldots, Q_N)$ (see Fig. 4). The situations in which V is composed of two or more subsequences possibly separated by central occlusions is very uncommon due to the nature of the images used by zoologists for (manual) photo identification, thus it can be ignored.

The aim of the global matching phase is to compare $(Q_1, \ldots, Q_N)$ with $(P_1, \ldots, P_M)$ looking for two subsequences of k nicks $(Q_j, Q_{j+1}, \ldots, Q_{j+k-1})$ and $(P_i, P_{i+1}, \ldots, P_{i+k-1})$ with similar positions with respect to the ear's reference points, respectively, $\mathbf{a}_I$, $\mathbf{b}_I$ and $\mathbf{a}_Q$, $\mathbf{b}_Q$ (see Section 3). While in this step we are interested only in comparing the *positions* of $Q_j, Q_{j+1}, \ldots, Q_{j+k-1}$ and $P_i, P_{i+1}, \ldots, P_{i+k-1}$, in Section 5 we show how the system compares the *shapes* of the two nick sets.

Note that if $M \leqslant N$ then V is delimited by either P_1 or P_M. This situation is shown in Figs. 4(a)–(c). Vice versa, if $M > N$ then V is delimited by either Q_1 or Q_N. The algorithm for the case in which $M \leqslant N$ is described below, while the case $M > N$ is dealt with analogously by exchanging $Q_1, \ldots, Q_N$ with $P_1, \ldots, P_M$.

Global Matching $(Q_1, \ldots, Q_N,\ P_1, \ldots, P_M)$
1 $\mu := \infty$.
*** Case P_1 is matched with Q_j $(j \in [1, N])$, see Figs. 4(a) and (b) ***
2 For each $j \in [1, N]$, do:
3 $k := \min\{M, N - j + 1\}$. *** This is the size of V ***
4 $ok :=$ *CheckPosition*$(P_1, \ldots, P_k, Q_j, \ldots, Q_{j+k-1}, k)$.
5 If ok, then:
6 $d :=$ *Shape Difference* $(P_1, \ldots, P_k, Q_j, \ldots, Q_{j+k-1}, k)$.
7 If $\mu > d$ then $\mu := d$.
*** Case P_1 cannot be matched with any Q_j, see Fig. 4(c) ***
8 For each $j \in [1, M-1]$, do:
9 $ok :=$*CheckPosition*$(P_{M-j+1}, \ldots, P_M, Q_1, \ldots, Q_j, j)$.
10 If ok, then:
11 $d :=$ *Shape Difference* $(P_{M-j+1}, \ldots, P_M, Q_1, \ldots, Q_j, j)$.
12 If $\mu > d$ then $\mu := d$.
13 Return μ.

Figs. 4(a)–(c) show V (the gray slice) delimited, respectively, by $P_1 = Q_j$ and P_M (a); $P_1 = Q_j$ and Q_N (b); and Q_1 and $P_M = Q_j$ (c). The first two cases are dealt with by the above algorithm in Steps 2–7, with, respectively, $k = M$ and $k = N - j + 1$, while the last case is dealt with in Steps 8–12. The function *CheckPosition*() is responsible to check the position consistency of the current ordered subsequence of matched nicks. This is done using the corresponding normalized endpoints (see Section 3) as follows:

CheckPosition $(P_1, \ldots, P_k, Q_1, \ldots, Q_k, k)$
1 $ok := true$.
2 For each $l \in [1, k]$ and while ok, do:
3 Let $\mathbf{e}'$, $\mathbf{f}'$ be the endpoints of P_l, normalized with respect to $\mathbf{a}_I$ and $\mathbf{b}_I$ and $\mathbf{g}'$, $\mathbf{h}'$ the endpoints of Q_l, normalized with respect to $\mathbf{a}_Q$ and $\mathbf{b}_Q$.
4 $d_1 := \|\mathbf{e}' - \mathbf{g}'\|_2$, $d_2 := \|\mathbf{f}' - \mathbf{h}'\|_2$.
5 $ok := d_1 \leqslant \lambda_1 \bigwedge d_2 \leqslant \lambda_1$.
6 Return ok.

In Step 5 above λ_1 is a threshold whose value must be chosen large enough in order to take into account that the endpoints of each nick are normalized using an approximate reference system. In fact, as mentioned in Section 3, the user selection of points **a** and **b** in each photo is a somewhat arbitrary operation. Note that, if both Q and I represent the same ear E, then the representations of both $\mathbf{e}'$, $\mathbf{f}'$ and $\mathbf{g}'$, $\mathbf{h}'$ are normalized with respect to a same reference system and the values d_1 and d_2 are invariant with respect to symmetric transformations.

If *CheckPosition*() returns *true*, then the current candidate sequence V of common visible nicks in I and Q is further analyzed in order to compute the shape dissimilarity among pairs of matched curves (see the next section). Once all the possible matches have been analyzed, the minimum computed dissimilarity value (μ) is returned in Step 13 of the *Global Matching* algorithm. This is the value which will be used to rank the current photo I with respect to the other database images. The resulting decreasing ordered sequence of database photos is finally shown to the user. Figs. 5 and 8 show the first four individuals ranked by the system in correspondence of the query shown in Fig. 2(c).

5. Shape comparison

For each subsequence $(P_1, \ldots, P_k)$ and $(Q_1, \ldots, Q_k)$ of nicks with a corresponding valid global position matched in the *Global Matching* algorithm, we need to verify the shape similarity of each pair of nicks (Q_i, P_i). Given a query (Q_i) and an image (P_i) nick, the shape difference of the corresponding curves can be estimated by subsampling Q_i and P_i using the same number of points (n) and then computing the squared Euclidean distance between pairs of corresponding points $(\mathbf{q}_j, \mathbf{p}_j)$, $\mathbf{q}_j \in Q_i$ and $\mathbf{p}_j \in P_i$ $(1 \leqslant j \leqslant n)$. However, before to compare Q_i and P_i, the two curves have to be aligned, i.e., they must be represented in a symmetric-invariant reference system in order to take into account for possible viewpoint changes. Nevertheless, since the shape comparison of two curves must be *carefully* estimated, we cannot use the reference points $\mathbf{a}_I$, $\mathbf{b}_I$, $\mathbf{a}_Q$ and $\mathbf{b}_Q$ for this purpose (see Sections 3 and 4). Rather, we use the nick-specific endpoints $\mathbf{e}_i$, $\mathbf{f}_i$, $\mathbf{g}_i$ and $\mathbf{h}_i$. Moreover, we have to take into account that either different users or the same user in different working sessions can select slightly different endpoints for the same nick in different photos of the same animal. For this reason we need to compare the nick defined by the endpoints **e** and **f** with the nick defined by a pair of endpoints chosen in the proximity of **g** and **h**. Of course, fixing **g** and **h** while varying **e** and **f** in a range around the user selected positions would produce the same or a very similar effect.

If $\mathbf{g}_i$ and $\mathbf{h}_i$ are the endpoints of Q_i and $\overline{Q}$ is the edge map of Q, then the neighborhoods $N(\mathbf{g}_i)$, $N(\mathbf{h}_i)$, respectively, of $\mathbf{g}_i$ and $\mathbf{h}_i$ are defined using:

$$N(\mathbf{x}) = \{\mathbf{p} \in \overline{Q} : \|\mathbf{x} - \mathbf{p}\|_2 \leqslant \lambda_2, \text{ and } \mathbf{x} \text{ and } \mathbf{p} \text{ are connected}\}, \quad (3)$$

where λ_2 is a distance threshold. We omit the algorithm for the effective construction of $N(\mathbf{g}_i)$ and $N(\mathbf{h}_i)$ since it is a trivial visit of the connected edge point set of $\overline{Q}$ starting from, respectively, $\mathbf{g}_i$ and $\mathbf{h}_i$. $N(\mathbf{g}_i)$ and $N(\mathbf{h}_i)$ are subsampled.

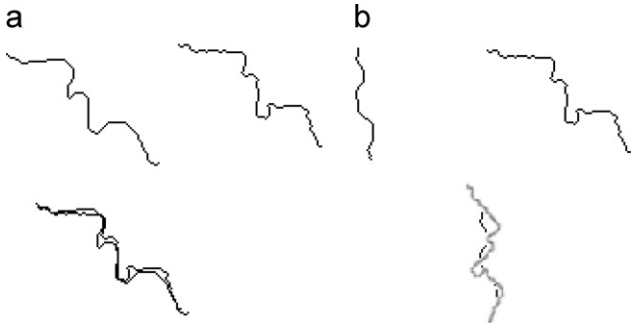

Fig. 7. A graphical representation of the nick curve comparison. (a) The query shape (top right curve) is compared with the nick of Fig. 6(b) (top left). (b) The query shape is compared with the nick of Fig. 6(c).

The below algorithm *Shape Difference* shows how $N(\mathbf{g}_i)$ and $N(\mathbf{h}_i)$ are used in order to hypothesize a set of reference systems able to align P_i with the nick Q_i suitably re-defined depending on different choices of the endpoints. For each pair of points $(\mathbf{q}_w, \mathbf{q}_z) \in N(\mathbf{g}_i) \times N(\mathbf{f}_i)$, let $C^{(\mathbf{q}_w, \mathbf{q}_z)}$ be the nick defined by the endpoints $\mathbf{q}_w$ and $\mathbf{q}_z$. The *Shape Difference* algorithm assumes that $\mathbf{q}_w$ corresponds to the endpoint $\mathbf{e}_i$ of P_i and $\mathbf{q}_z$ corresponds to $\mathbf{f}_i$. The goodness of this hypothesis must be checked by computing the shape difference between $C^{(\mathbf{q}_w, \mathbf{q}_z)}$ and P_i in a symmetric-invariant fashion. For this reason we normalize the points of $C^{(\mathbf{q}_w, \mathbf{q}_z)}$ using Eq. (2) and reference endpoints $\mathbf{q}_w$ and $\mathbf{q}_z$ in place of **e** and **f** (see Section 3). For saving computational time, the result is stored in $Q'_i[w, z]$ so that it can be re-used in comparing the query Q with different database images. We can now compute the point-to-point difference between Q_i and P_i and iterate the process for a different choice of endpoints $\mathbf{q}_w$ and $\mathbf{q}_z$. The details of the algorithm are shown below:

Query Nick Normalization $(Q_1, \ldots, Q_N)$
1 For each $i \in [1, N]$, do:
2 For each $(\mathbf{q}_w, \mathbf{q}_z) \in N(\mathbf{g}_i) \times N(\mathbf{h}_i)$, do:
3 Let C be the shortest path between $\mathbf{q}_w$ and $\mathbf{q}_z$ in $\overline{Q}$.
4 $Q_i^{(\mathbf{q}_w, \mathbf{q}_z)} :=$ *Subsampling Procedure* (C).
5 For each point $\mathbf{q}_j \in Q_i^{(\mathbf{q}_w, \mathbf{q}_z)}$, do:
6 Let $\mathbf{q}'_j$ be the corresponding normalized point obtained using Eq. (2) and reference endpoints $\mathbf{q}_w$ and $\mathbf{q}_z$.
7 $Q'_i[w, z, j] := \mathbf{q}'_j$.

The existence of a connected path between $\mathbf{q}_w$ and $\mathbf{q}_z$ in Step 3 is guaranteed by the construction of Q_i, $N(\mathbf{g}_i)$ and $N(\mathbf{h}_i)$. C is then subsampled using the algorithm presented in Section 3. The resulting curve $Q_i^{(\mathbf{q}_w, \mathbf{q}_z)}$ is represented by the same number (n) of points used to represent all the image nicks. For this reason in the below algorithm *Shape Difference* we can directly compare the jth point $(\mathbf{q}'_j)$ of $Q'_i[w, z]$ with the jth point $(\mathbf{p}'_j)$ of P_i. This is done in Step 6 where the difference between P_i and $Q'_i[w, z]$ is stored in $d^{(\mathbf{q}_w, \mathbf{q}_z)}$. Since we want that the difference between P_i and Q_i is given by the best possible alignment between the two nicks, the whole process is iterated

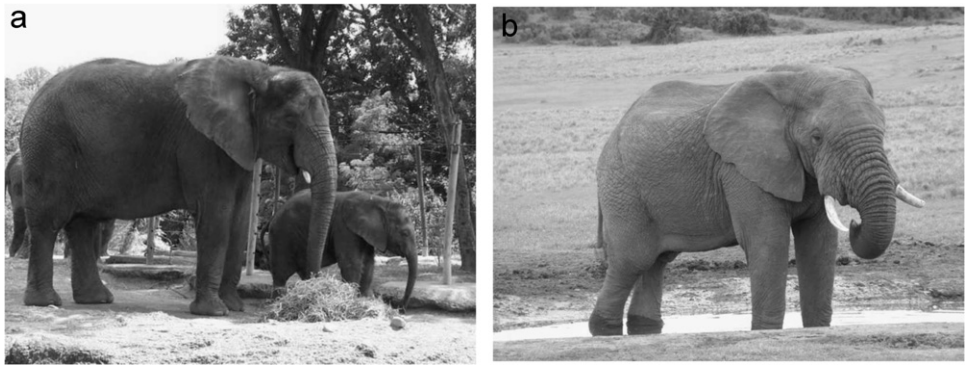

Fig. 8. The third (a) and the fourth (b) individual ranked by the system with respect to the query of Fig. 2(c).

for every pair of possible endpoints $(\mathbf{q}_w, \mathbf{q}_z) \in N(\mathbf{g}_i) \times N(\mathbf{h}_i)$ and the minimum value of $d^{(\mathbf{q}_w,\mathbf{q}_z)}$ is taken as the shape dissimilarity value between P_i and Q_i. Fig. 7(a) shows a graphical representation of the shape dissimilarity between the nick extracted from the query image of Fig. 3(a) and a nick of a database image corresponding to the same individual as the query, while Fig. 7(b) shows the comparison of the same query with a nick of a different animal.

Shape Difference $(P_1, \ldots P_k, Q_1, \ldots, Q_k, k)$
1 $d := 0$.
2 For each $i \in [1, k]$, do:
3 $d_i := \infty$.
4 For each $(\mathbf{q}_w, \mathbf{q}_z) \in N(\mathbf{g}_i) \times N(\mathbf{h}_i)$, do:
5 Let $P'_i = \{\mathbf{p}'_1, \ldots, \mathbf{p}'_n\}$ be the ith nick normalized w.r.t. $\mathbf{e}_i$ and $\mathbf{f}_i$.
6 $d^{(\mathbf{q}_w,\mathbf{q}_z)} := \sum_{j=1}^{n} \|\mathbf{p}'_j - Q'_i[w, z, j]\|_2^2$.
7 If $d_i > d^{(\mathbf{q}_w,\mathbf{q}_z)}$ then $d_i := d^{(\mathbf{q}_w,\mathbf{q}_z)}$.
8 $d := d + d_i$.
9 $d := d/k$.
10 Return d.

Finally, the above described curve comparison procedure is iterated for every pair of curves (P_i, Q_i) previously matched by the *Global Matching* algorithm (Section 4) and the average dissimilarity value d is returned in Step 10. Note that, since every curve $Q'_i[w, z]$ is subsampled using the same number of points n, we do not need to assign different weights to each partial difference d_i in the sum of Step 8, nor we need to normalize the sum in Step 6 with respect to the number of compared points n.

6. Discussion

The computational complexity of the *Global Matching* algorithm is trivially given by $O((N + M - 1)(p + s))$, where $p = O(k)$ is the complexity of *CheckPosition* and s is the complexity of *Shape Difference*. This latter in turn is given by $s = O(km^2n)$, where $m = \max\{\sharp N(\mathbf{g}_i), \sharp N(\mathbf{h}_i)\}$, being $\sharp X$ the cardinality of the set X, while $n = \sharp P'_i$. Since $M \leqslant N$ (see Section 4) and $k \leqslant N$, the whole matching worst case computational complexity is upper bounded by

$$O(DN^2m^2n), \tag{4}$$

where D is the number of photos stored in the system's database. Typical magnitude orders for the factors involved in Eq. (4) for a medium size zoological database are: a few hundreds of individuals for D, one to five nicks per ear (N), a pair of dozens of points for n and the same for m.

We want to point out that searching for a pair of points $(\mathbf{q}_w, \mathbf{q}_z) \in N(\mathbf{g}_i) \times N(\mathbf{h}_i)$ to match with $(\mathbf{e}_i, \mathbf{f}_i)$ and then comparing P'_i with $Q'_i[w, z]$ (respectively normalized with respect to $(\mathbf{e}_i, \mathbf{f}_i)$ and $(\mathbf{q}_w, \mathbf{q}_z)$) is conceptually equivalent to search for a symmetric 2D transformation T mapping $\overline{Q}$ upon $\overline{I}$ in such a way to overlap the points $\mathbf{q}_w$ upon $\mathbf{e}_i$ and $\mathbf{q}_z$ upon $\mathbf{f}_i$. The latter is the approach which common *model based* object recognition methods are based on Refs. [19–22]. However, the shape dissimilarity value computed in Step 6 of the *Shape Difference* algorithm between P'_i and $Q'_i[w, z]$ is much more accurate than common model based shape verification techniques. In fact, we not only check that the generic point $\mathbf{q}'_j$, once projected into $\overline{I}$ is sufficiently close to *any* point $\mathbf{p}'$ of P'_i (or of $\overline{I}$). We also impose that the distance must be computed exactly with the jth point $\mathbf{p}'_j$ of P'_i. This ordering implicitly introduces context information in the point matching procedure. Conversely, common methods for model-based shape verification are *context-free*, which in images with thick textures and cluttered backgrounds can lead to a high number of false positives (e.g., see [21, p. 416–418]).

An example is given by the *chamfer distance*-based matching algorithms [23,24] which, given an edge map $\overline{I}$ of $W \times H$ pixels, before to run the actual matching phase pre-compute a corresponding map $C_{W\times H}$. $C(\mathbf{p}) = d$ indicates that there exists a point $\mathbf{p}_1 \in \overline{I}$ such that $\|\mathbf{p} - \mathbf{p}_1\|_x = d$ and for all the other points $\mathbf{p}_2 \in \overline{I}$, $\|\mathbf{p} - \mathbf{p}_2\|_x \geqslant d$. Usually the Norm used is L_1 ($\|\cdot\|_1$). Once the model of the searched for shape M is projected into $\overline{I}$ in a given supposed position $T(M)$, the verification of the goodness of such an alignment hypothesis T can be performed by iteratively computing, for each $\mathbf{q} \in M$ the value of $C(T(\mathbf{q}))$. A very similar class of methods are those based on the Hausdorff distance [25]. Nevertheless, $C(\mathbf{p})$ does not contain any context-dependent information. In other words, given a connected line segment $(\mathbf{q}_1, \ldots, \mathbf{q}_m)$ contained in M and supposing that, for each $i \in [1, m]$, $\mathbf{p}_i \in \overline{I}$ is the closest point to $T(\mathbf{q}_i)$, then there is no guarantee that the points $\mathbf{p}_1, \ldots, \mathbf{p}_m$

belong to the same object neither that they are connected in $\overline{I}$, even if $\sum_{i=1}^{m} C(T(\mathbf{q}_i))$ is a small value. For instance, in images with regions dense of edge pixels (e.g., Fig. 3) it is likely that $\sum_{i}^{m} C(T(\mathbf{q}_i))$ is small due to sets of points $\{\mathbf{p}_1, \ldots, \mathbf{p}_m\}$ randomly clustered.

A similar problem due to the same reasons is present in Hough-like approaches. In fact, it is well known that one of the most important problems with Hough-like techniques are the false positives in images with thick textured regions (e.g., see [21, p. 330]). Indeed, the Hough accumulator H represents all the possible parameter discrete values of the searched for geometric transformation T of the shape model into the image reference system. Nevertheless, uniformly distributed edge pixels in $\overline{I}$ can randomly concentrate their votes in few H's buckets producing high peaks not actually corresponding to any significant shape in the image [21,26]. As mentioned in Section 2, we have implemented the traditional Ballard's GHT and tested it in our elephant domain in which edge images are usually very cluttered. The experimental results confirmed the inadequacy of such a technique due to the high number of false positives.

Conversely, in our proposed matching method possible surrounding lines and/or isolated edge pixels of $\overline{I}$ close to P_i' are not taken into account by the shape comparison process and the jth point of the curve Q_i is compared with only the jth point of the curve P_i'. This is possible thank to both a minor user intervention in selecting approximate landmark points (e.g., $\mathbf{e}_i$, $\mathbf{f}_i$) in each image and a subsequent hypothesis-and-test process performed in the algorithms *Global Matching* and *Shape Difference* in order to refine such selections.

The method we propose does not depend on specific assumptions of the elephant domain and the curves representing interesting nicks which are compared in the *Shape Difference* algorithm have no restriction on the possible shape. For these reasons the multi-curve matching approach presented in this article can be used in domains different from animal photo identification. Possible applications are those domains in which *Content Based Image Retrieval* (CBIR) can be performed using the shape of one or more details of a given object. Such details do not need to be connected and an approximate user selection of candidate positions in each image to analyze allows the system to perform a reliable detection also in very cluttered and noisy images.

7. Experimental results

For testing purposes we have used an elephant photo database provided by the Zoological Department of the University of Rome "La Sapienza", which represents individuals of the elephant population of Zakouma Ciad National Park [1]. Almost all the photos are actually frames extracted from a few (low-resolution) video sequences and previously used by the researchers of the same department for manual elephant identification. All the images represent elephants in a profile pose but with unknown positions with respect to the camera and possible minor ear occlusions (see Fig. 1 for a few examples). The database used for testing is composed of 332 images of 268 different individuals. Each individual is represented by one to four photos, respectively, showing either the left or the right ear, and each ear of the image repository contains from one to four nicks. Other 200 different photos of the same individuals of the database have been randomly selected and used as queries. The photos used as queries have not been included into the system's database. The 200 query photos correspond to 200 out of the 268 individuals of the database.

The experiments have been performed by zoologist researchers not involved in the system's implementation. Table 1 shows the results of their trials. The second row shows the number of queries for which the corresponding correct individual of the database has been ranked in the ith position. The last column refers to those queries for which no match has been found. The reason for this is that some of the queries (24, i.e., 12% of the total) have been wrongly rejected by the *Global Matching* algorithm (see Section 4) mainly due to a mismatch in the selection of the reference points (**a** and **b**). The last row of Table 1 shows the *Cumulative Matching Characteristic* (CMC), defined as the probability $CMC(r)$ that, for a given query, the correct corresponding individual of the database is classified by the system in the first r positions. The CMC values are often graphically shown using *boxplots* like the ones in Fig. 9. The box in a boxplot shows the portion of the data (number of test queries in our case) ranging from 25% to 75% of the admissible values (position of the correct individual in our case). Fig. 9 and Table 1 show that in about 50% of the queries the correct individual has been retrieved in the very first position and that in about 75% of the trials the user needed to scan only the first five photos shown by the system. Moreover, taking into account only the query photos not rejected by the *Global Matching* algorithm, we have that almost all of them, 162 out of 176 (92%) have been correctly ranked in the first six positions, being the remaining 14 items sparsely spread on the other possible positions. This fact underlines the effectiveness of the shape comparison based on the techniques proposed in Section 5.

An important issue in CBIR system's evaluation is the dependency on the test database cardinality. In Ref. [27] Huijsmans and Sebe show how precisions and recall of a CBIR system usually strongly degrade by augmenting the number of items in the database which are different from the searched for class. In order to relate our results with our test database size, we adopt some suggestions proposed in Ref. [27], where the authors indicate with $g = c/d$ the *generality* of a class of the database, i.e., the number of objects of that class (or "relevant items", c) with respect to the size of the whole database (d). In our case we have different classes, being each class a different individual of the database, each of which is represented by one to four photos. However, the relative positions of points **a** and **b** allows the system to automatically select the right/left profile (Section 3) hence discarding about an half of the database photos. Thus, we set $c = 1.12$, which is the average number of photos per individual per profile in the database. The size of the repository is: $d = 332/2 = 166$, thus $g = 0.007$. This value represents the *expected random retrieval rate*, i.e. the expected number of true positives which would be obtained by a random output system in the very first position of the ranked list [27].

Table 1
Number of queries for which the corresponding correct individual of the database has been ranked by the system in the first 10 positions

Position	1	2	3	4	5	6	7	8	9	10	N.C.
Query ♯	92	32	5	13	4	16	0	1	4	3	24
CMC	0.46	0.62	0.64	0.71	0.73	0.81	0.81	0.81	0.83	0.85	–

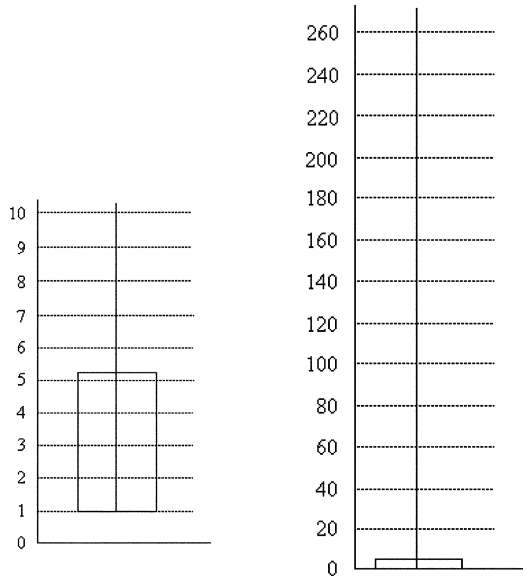

Fig. 9. The boxplots corresponding to Table 1. (a) CMC values with respect to the first 10 individuals ranked by the system. (b) CMC values with respect to all the individuals of the database. The vertical line ranging for all the database individuals indicates that in a few cases (24), no match at all has been found.

Comparing g with the probability (v) to obtain a relevant item in the first position using our proposed system (i.e., $CMC(1)$: see Table 1), we have $v = 0.46$ against $g = 0.007$.

Preliminary results, obtained with a subset (about 100 images of 60 individuals) of the same database [1] produced similar success percentages ($v = 0.5$). However, database sizes considerably greater than the present could likely lower down the precision of the system due to the frequency of individuals with nicks with similar shapes and similar positions which make hard the recognition task also for a human being. Nevertheless, typical zoologist photo databases are not so huge, hence our system can be effective for medium size archives (a few hundreds of photos).

To the best of our knowledge, no other elephant photo identification system has been proposed so far, hence a direct comparison is difficult, while single curve matching-based approaches would not work in our domain for the reasons discussed in Section 2. However, the single curve matching approach proposed in Ref. [7] for marine mammals photo identification achieves worse results than our method. For instance, in the sea lion database (92 images of 37 individuals), in about 50% of the queries the proposed system ranks the correct individual in the first 7 positions and in about 75% of the queries in the first 28 positions. Similar values are obtained in the gray whale and in the dolphin databases, with, respectively, 95 images of 37 individuals and 75% of the queries in the first 23 positions and 624 images of 164 individuals and 75% of the queries in the first 30 positions.

In Ref. [28] two different Sperm Whale photo identification systems are compared using a database of 592 photos of 296 different individuals and 150 query photos. The two compared systems are: Europhlukes [8], based on a wavelet transform of the fluke contour's representation and Highlight, in which the user is requested to input the number of nicks, scallops and waves present on the animal's tail [29]. Both methods have obtained less than 78% of relevant items in the first 9 positions (against the 83% obtained by our system). Using a subset of the system composed of only high-quality photos highlight obtained 87.6% of correct matches in the first 9 positions and Europhlukes 86%, slightly outperforming our system (tested with low-quality photos). However it should be noted that Europhlukes is based on the possibility to extract the whole connected curve representing the tail's characterizing shape (see Sections 1 and 2) and hence cannot be realistically used in our domain in which the elephant's ear border usually has insufficient contrast with respect to the body skin which makes the extraction of a connected curve particularly hard (see Section 2). Conversely, Highlight needs a human intervention in the nick type classification, making the system less automatic and more dependent on the human subjective judgment and expertise [28].

8. Conclusions

We have presented an elephant photo identification system based on the comparison of the shape details of nick subsequences characterizing the animals' ears. The proposed method is not dependent on specific assumptions on the elephant domain and the photos used are generally low resolution and highly cluttered images. Segmentation and recognition are performed using as basis rough position information input by the user. Starting from the user suggested initial positions a set of normalization hypotheses is formulated by the system in order to align the query image shape details with the curves stored in the system's database. The achieved recognition accuracy outperforms similar single curve matching-based animal recognition approaches when tested in low-medium quality images and allows the user to find the correct individual for a given query with a high probability in the very first positions ranked by the system.

The present version of our system is currently used by researchers of the Zoological Department of the University of Rome "La Sapienza" for elephant identification. We are in contact with zoologists of the Centre for African Conservation

Ecology (South Africa) for testing an improved version of the system with their databases.

Acknowledgments

This work has been done in collaboration with the Zoological Department of the University of Rome "La Sapienza". A particular thank to Francesca Della Rocca, who suggested us the idea of an automatic system for elephant photo identification and gave us the know how about elephant (manual) photo recognition as well as the database of images used in the experiments. We also want to thank Luciana Massaro, Dafne Petriccione and Sabrina Rossi for their passion, their patience and their precious help in testing the system.

References

[1] D. Petriccione, A. Ardovini, L. Cinque, F. Della Rocca, L. Massaro, A. Ortolani, E. Sangineto, Computer-aided wild elephant identification, in: 11th Congress of the International Society for Behavior Ecology, 2006, p. 264.
[2] E. Ranguelova, M. Huiskes, E.J. Pauwels, Towards computer-assisted photo-identification of humpback whales, in: Proceedings of the IEEE International Conference on Image Processing, 2004, pp. 1727–1730.
[3] E. Ranguelova, E.J. Pauwels, Saliency detection and matching strategy for photo-identification of humpback whales, in: ICGST International Journal on Graphics, Vision and Image Processing, 2006 (Special Issue on Features and Analysis).
[4] C. Gope, N. Kehtarnavaz, G. Hillman, Zernike moment invariants based photo-identification using Fisher discriminant model, in: Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2004, pp. 1455–1458.
[5] B.N. Araabi, N. Kehtarnavaz, A string matching computer-assisted system for dolphin photoidentification, Biomed. Eng. 28 (2000) 1269–1279.
[6] B.B. Kimia, Shape representation for image retrieval, Image Databases: Search and Retrieval of Digital Imagery, 2002, pp. 345–372.
[7] C. Gope, N. Kehtarnavaz, G. Hillman, B. Wursig, An affine invariant curve matching method for photo-identification of marine mammals, Pattern Recognition 38 (2005) 125–132.
[8] R. Huele, H.A. Udo de Haes, J.N. Ciano, J. Gordon, Finding similar trailing edges in large collections of photographs of Sperm Whales, J. Cetacean Res. Manage. 2 (3) (2000) 173–176.
[9] F. Mokhtarian, Silhouette-based isolated object recognition through curvature scale-space, IEEE Trans. PAMI 17 (5) (1995) 539–544.
[10] F. Mokhtarian, S. Abbasi, Shape similarity retrieval under affine transforms, Pattern Recognition 35 (1) (2002) 31–41.
[11] H. Tek, B.B. Kimia, Boundary smoothing via symmetry transforms, J. Math. Imaging Vision 14 (3) (2001) 211–223.
[12] A. Folkers, H. Samet, Content-based image retrieval using Fourier descriptors on a logo database, in: ICPR '02, vol. 3, 2002, pp. 521–524.
[13] M. Kass, A. Witkin, D. Terzopoulos, Snakes: active contour models, Int. J. Comput. Vision 1 (1988) 321–331.
[14] Z. Xue, S.Z. Li, E.K. Teoh, AI-EigenSnake: an affine-invariant deformable contour model for object matching, Image Vision Comput. 20 (2) (2002) 77–84.
[15] T.F. Cootes, C.J. Taylor, D.H. Cooper, J. Graham, Training models of shape from sets of examples, in: Proceedings of the British Machine Vision Conference, 1992, pp. 9–18.
[16] D.H. Ballard, Generalizing the Hough transform to detect arbitrary shapes, Pattern Recognition 13 (2) (1981) 111–122.
[17] A.A. Kassim, T. Tan, K.H. Tan, A comparative study of efficient generalized Hough transform techniques, Image Vision Comput. 17 (10) (1999) 737–748.
[18] J. Canny, A computational approach to edge detection, IEEE Trans. PAMI 8 (6) (1986) 679–698.
[19] D.P. Huttenlocher, S. Ullman, Recognizing solid objects by alignment with an image, Int. J. Comput. Vision 5 (2) (1990) 195–212.
[20] N. Ayache, O.D. Faugeras, HYPER: A new approach for the recognition and positioning of two-dimensional objects, IEEE Trans. PAMI 8 (1) (1986) 44–54.
[21] D.A. Forsyth, J. Ponce, Computer Vision: A Modern Approach, Prentice-Hall, Englewood Cliffs, NJ, 14 August 2003 ISBN: 0130851981.
[22] S. Ullman, High-level vision. Object recognition and visual cognition, in: A Bradford Book, The MIT Press, Cambridge, MA, 1996.
[23] C. Fouard, G. Malandain, 3-D chamfer distances and norms in anisotropic grids, Image Vision Comput. 23 (2) (2005) 143–158.
[24] A. Thayananthan, B. Stenger, P.H.S. Torr, R. Cipolla, Shape context and Chamfer matching in cluttered scenes, in: CVPR '05, vol. 1, 2003.
[25] D.P. Huttenlocher, G.A. Klanderman, W.J. Rucklidge, Comparing images using Hausdorff distance, IEEE Trans. PAMI 15 (9) (1993) 850–863.
[26] M. Anelli, A. Micarelli, E. Sangineto, A new content based image retrieval method based on a sketch-driven interpretation of line segments, in: Proceedings of IJCAI-2003, Acapulco, Mexico, 2003.
[27] D.P. Huijsmans, N. Sebe, How to complete performance graphs in content-based image retrieval: add generality and normalize scope, IEEE Trans. PAMI 27 (2) (2005) 245–251.
[28] B.W.P.M. Beekmans, H. Whitehead, R. Huele, L. Steiner, A.G. Steenbeek, Comparison of two computer-assisted photo-identification methods applied to Sperm Whales (*Physeter macrocephalus*), Aquat. Mamm. 31 (2) (2005) 243–247.
[29] H. Whitehead, Computer-assisted individual identification of Sperm Whales flukes, Reports of the International Whaling Commission vol. 12, 1990, pp. 71–77 (special issue).

About the Author—ALESSANDRO ARDOVINI received the degree in Computer Science from the University of Rome "La Sapienza" in 2006. He is currently working as software developer at Systemies s.r.l.

About the Author—LUIGI CINQUE received his degree in Physics from University of Napoli in 1983. After a few years spent at Artificial Intelligence Laboratory (Selenia S.p.A), working on the development of expert systems and knowledge-based vision systems, in 1990 he joined the Department of Computer Science of the University "La Sapienza" of Rome as Assistant Professor. Presently, he is Full Professor of Computer Science.
His current interests include computer vision, parallel image analysis and architectures, shape and object recognition, image sequences analysis, and image database.
He is a Senior Member of IEEE Computer Society, a member of the Association for Computing Machinery, and a member the Pattern Recognition Society.

About the Author—ENVER SANGINETO received his degree in Computer Science at the University of Pisa in 1995. In 2001 he received a Ph.D. in Computer Engineering at the University of Rome "La Sapienza", discussing a thesis on: "Object Classification through Geometric Abstraction". He has been working from 2001 to 2006 at the Artificial Intelligence Laboratory of the University of Rome "Roma Tre" and with the CRMPA (a research centre of the University of Salerno, Italy). In 2006 he joined the Department of Computer Science of the University "La Sapienza" of Rome with a post-Doct. position. His main scientific interests concern computer vision, pattern recognition, e-learning, and robotics. He has published various papers on these topics in important national and international conferences and journals.