EURASIP Journal on Advances in Signal Processing
a SpringerOpen Journal

RESEARCH Open Access

Hand posture recognition using jointly optical flow and dimensionality reduction

Nabil Boughnim, Julien Marot, Caroline Fossati and Salah Bourennane*

Abstract

Hand posture recognition is generally addressed by using either YC_bC_r (luminance and chrominance components) or *HSV* (hue, saturation, value) mappings which assume that a hand can be distinguished from the background from some colorfulness and luminance properties. This can hardly be used when a dark hand, or a hand of any color, is under study. In addition, existing recognition processes rely on descriptors or geometric shapes which can be reliable; this comes at the expense of an increased computational complexity. To cope with these drawbacks, this paper proposes a four-step method recognition technique consisting of (i) a pyramidal optical flow for the detection of large movements and hence determine the region of interest containing the expected hand, (ii) a preprocessing step to compute the hand contour while ensuring geometric and illumination invariance, (iii) an image scanning method providing a signature which characterizes non-star-shaped contours with a one-pixel precision, and (iv) a posture classification method where a sphericity criterion preselects a set of candidate postures, principal component analysis reduces the dimensionality of the data, and Mahalanobis distance is used as a criterion to identify the hand posture in any test image. The proposed technique has been assessed in terms of its performances including the computational complexity using both visual and statistical results.

Keywords: Hand posture; Contour signature; Classification algorithm; Optical flow; Principal component analysis

1 Introduction

Hand gesture and posture classification are of increasing interest for human-computer interaction. Previous works have concentrated on hand gesture classification (see [1,2] and references in [3,4]) where gesture command is based on slow movements with large amplitudes. Movements of the entire body were also investigated [5] for the purpose of action classification using of 3D representation and a 3D thinning algorithm. We believe that future applications should consider the classification of planar hand postures. This task is difficult because each finger must be distinguished and, if a user wishes to afford a large dictionary of postures, some of them may be similar. The common approach for hand posture characterization is based on descriptors like Hu moments [6], Zernike moments [7], and Fourier descriptors [8,9]. The advantages of these methods relate to their intrinsic geometric invariance (i.e., translation, rotation, and scaling). However, the results given in [10], and also presented in [3], show that Fourier descriptors do not take into account postures which are visually close. This is due to the low number of coefficients required to afford a moderate computational load: contours are smoothed and some details of the hand contours are skipped. Moreover, to the best of our knowledge, most hand posture recognition methods include a YC_bC_r mapping, which solely enhances hands with a color which is close to the white color. Consequently, they fail to characterize hands of colored people, or hands wearing colored gloves.

In this paper we propose a hand posture recognition method which overcomes the main drawbacks of existing methods [1,6,10]. Firstly, our method works independently of the hand color. To achieve this, we propose to adapt the optical flow for contour segmentation independently from its color where the hand contour is extracted from a color image. Secondly, we have improved the recognition performances especially for very similar postures by proposing an approach for hand posture characterization, which involves an original signature developed in [3] and improved in [4]. Thirdly, we have been able to

*Correspondence: salah.bourennane@fresnel.fr
CNRS-UMR 7249 / Fresnel Institute - Ecole Centrale Marseille, Aix Marseille Université, D. U. de Saint-Jérôme, 13397 Marseille Cedex 20, France

Springer

© 2013 Boughnim et al.; licensee Springer. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

keep the computational complexity including the memory requirements as low as possible. This has been possible by adapting a sphericity criterion to reduce the set of candidate postures and by reducing the dimensionality of the data.

The remainder of the paper is organized as follows: in Section 2 we discuss the working conditions of the system before giving the problem statement and the objectives. In Section 3 we explain how optical flow is adapted as a hand contour detection method and how adequate preprocessing methods yield a binary image containing only the hand contour. This allows the hand recognition process to be invariant to translation, scaling, and rotation. In Section 4 we provide details and interpretations of a proposed signature for hand posture characterization which improves slightly our seminal work [3]. In Section 5 we present a distance criterion which allows us to classify hand posture images; this distance criterion is associated with a sphericity criterion which is able to reduce the size of the dictionary of candidate postures. Also, a data dimensionality reduction algorithm is proposed to speed up the computation of this distance. In Section 6 we present a summary of the proposed method for hand posture recognition and a theoretical study of the computational load. The performances of the proposed algorithm are presented in Section 7, as well as the comparative results. Finally we draw conclusions in the light of these results and state some prospects in Section 8.

2 Problem statement and image acquisition setup

This section aims to state the problem of hand posture recognition and describes the dictionary of hand postures chosen in collaboration with an industrial partner. The practical setup dedicated to the acquisition of hand images is also discussed.

2.1 Problem statement

Systems which employ hand-driven human-machine interfaces (HMI) interpret hand gestures and postures in different modes of interaction depending on the application domain. Previous works have concentrated on hand gesture classification [1,2,10]. In [1,10], gesture command is based on slow movements with large amplitudes (see for instance in [1] the 12 types of hand gesture). In [2], local orientation histograms are computed. Very high recognition rates are obtained but only in simplified conditions (uniform background, limited alphabet) in order to facilitate the hand region extraction.

To the best of our knowledge, applications should consider more complex operation conditions for the purpose of automated sign language decoding for instance, or touchless technology for retail applications as food and beverage vending machines. As such, the underlying image processing algorithms should face rather complex environments. Moreover, hand gestures only are not sufficient to face the requirements of such applications. Unlike hand gestures, hand postures describe the hand shape and not its movement, and can be very different from each other. It is extremely difficult to recognize all possible configurations of the hand starting from its projection on a 2-D image. Indeed, some parts of the hand can be hidden.

There exist some reference databases of specific hand postures, such as the Triesch database [11], available on the Web [12]. However, this database has a number of limitations: the number of images is low, the viewing angle and the size and the orientation of the hand are always the same, the images are in gray level and contain only the hand, and the background is rather simple. In this work we consider an application where our industrial partner wishes to have a rather elevated number of postures, which can yield later to various instructions for a HMI: this dictionary of postures is meant for automatic distributors such as drink vending machines. For this reason, we have built our own database (also partially used in [3]). It contains 11 postures, which are displayed in Figure 1, with the corresponding posture index above. As will be shown in the results section, the background in these images is rather complex so that if our method works well with our database, it will certainly work with Triesch database.

These postures have been chosen to be easily performed by any person. They differ from the sign language which aims at easing lip reading. To afford a large dictionary, and thereby numerous functionalities of the human-machine interface, the dictionary must include a number of postures which is large enough. This implies that similar postures may be present in the dictionary. Some postures permit to test the discrimination performances of the proposed and comparative methods: they are visually very close, such as postures 4 and 5, as well as postures 8 and 9. The images of the database were obtained as follows: an expert user shoots a movie containing the 11 postures. Then the frames of the video are split to get the images in the RGB color space. These images compound the learning set. Other users which are not the expert user shoot a movie containing the 11 postures. The same process as for the learning database is adopted to get a set of images; these images compound the test set. To generate the learning database and the test set, a simple setup involving a camera is used.

2.2 Hand image acquisition setup

This setup contains a complementary metal oxide semiconductor (CMOS) camera, presented in Figure 2 which displays a side view in Figure 2a, a front view in Figure 2b, and a view from behind in Figure 2c. It has the size of a webcam and could further be integrated in an embedded system. The camera is placed over the desk surface, its axis is orthogonal to the desk surface. Wide angle optics

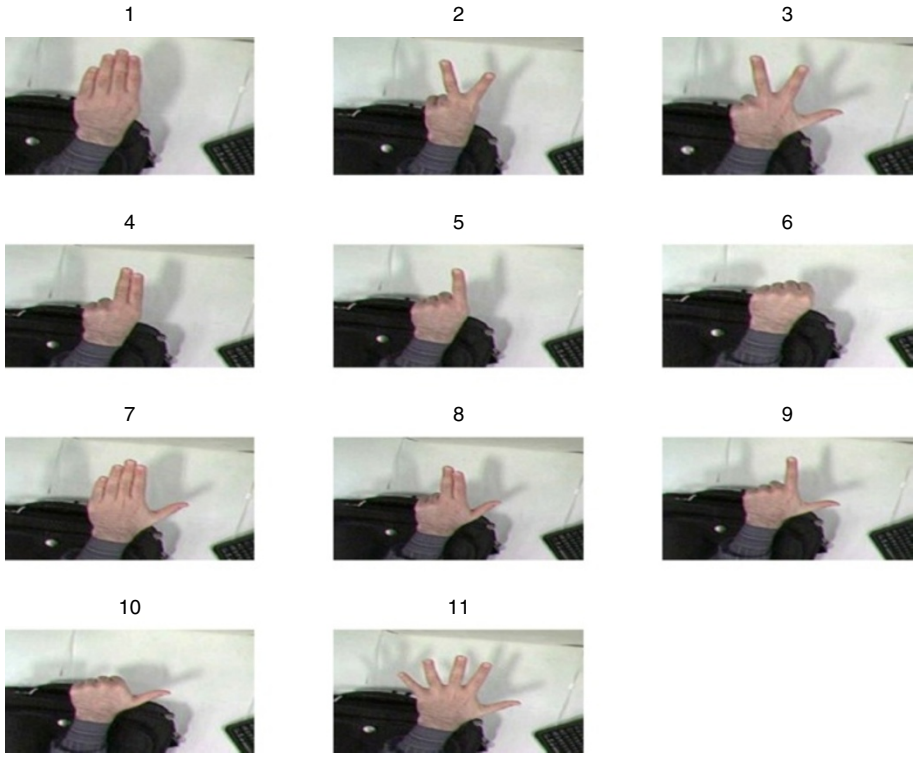

Figure 1 Postures of the database-cropped images. Each posture is placed below its corresponding index.

(90°) are used so that the field of vision is wide enough. The acquisition format can be either CIF or VGA. The video stream is transmitted to the computer by a USB connection in RGB format. The user can then interact with his computer and follow the evolution of his experiment directly on the screen.

The video stream is split into several frames, which are color images in RGB format. Each frame can then be processed, the first step being contour detection.

3 Detection and segmentation of the hand

In this paper, the term contour detection denotes the following operation: from an RGB frame provided by the camera, we aim at getting a binary image containing only the hand contour, as 1-valued pixels over a background of 0-valued pixels. This detection step is required for the next step, presented in a further section, which is contour characterization.

3.1 Overview of classical detection methods

A rule of thumb about contour characterization methods such as Fourier descriptors [10,13] is that they require a binary image I, possibly noise-free. The same constraint holds in the frame of our work. To perform hand contour detection, some classical preprocessing methods have been applied in previous works [3,4,10,13]: the YC_bC_r mapping and the selection of the C_r component emphasize the hand surface with respect to the background. Another mapping could also be used: in [14], HSV mapping is combined with a hue setting to distinguish the hand from other body parts. The results presented are convincing but this method exhibits a limitation: the user *must* wear a red glove. To our knowledge, no study based on mappings from a color space to another such as YC_bC_r or HSV provides good detection results on databases where the object of interest may exhibit any color. We want to develop a new method, which detects the hand independently of its color.

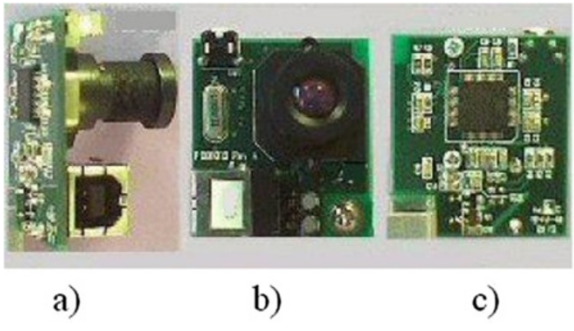

Figure 2 Camera. (a) Side view. **(b)** Front view. **(c)** View from behind.

In [15], Soriano et al. propose a dynamic skin color model, for a segmentation purpose. Their method copes with changes in illumination. However, their method still relies on specific color properties of the skin. No result is presented concerning dark skins or hands wearing color gloves. In [16], Raja et al. modelled their object colors as a Gaussian mixture and estimated its parameters. In [17], Yoo et al. segment faces by computing a chromatic histogram, which exhibits the advantage of being insensitive to scaling and rotation. However, Yoo et al. must combine the chromatic histogram with the prior knowledge of the approximate shape of faces to detect them. We wish to perform hand posture recognition, so we cannot assume any prior knowledge about the shape of the hand. In other papers dealing with hand gesture recognition, such as [18,19], good results are obtained on white hands but no result on black or colored hands is presented.

The main drawback of the previously cited methods is that they do not handle the hands of colored people or those wearing colored gloves. More generally, as pointed out in [20], color has been widely used for hand segmentation and many approaches rely on predefined color models. However, it is difficult to predefine a color model in an application where there is no control over the background, which may be of any color and complexity. It is also not sufficient if the color of an object on the background is very similar to the color of the hand. We propose to adapt the concept of optical flow for the detection of hands through their movement.

3.2 Optical flow as hand detector

This work proposes, for the first time, to adapt optical flow as a contour detection solution. Indeed, the sign language or touchless technology applications to which our posture recognition method is dedicated assume that the hand user is, at least, slowly moving. This is also coherent with the current way to use touchless technology.

Optical flow was first introduced to characterize movements between two frames [21-23], without any prior knowledge about the contents of the frames by determining the gray level values which have migrated from one point to another. The following sections are organized as follows: we firstly emphasize that the original work of [21] is based on a variational method upon which every variant of optical flow is still based and that some limitations were tackled by various successive modifications. Our choice is justified from a pyramidal version of optical flow [24]. Secondly, we focus on the applications of optical flow and we discuss the practical implementation of the proposed adaptation of optical flow with a view to propose a method to improve it.

The principles of optical flow are as follows: one can associate some kind of velocity with each pixel in the frame or, equivalently, some distance a pixel has covered between the previous frame and the current one. Such a construction is usually referred to as a dense optical flow [21], which associates a velocity with every pixel in an image. Horn and Schunk, in [21], introduced a novel framework where the optical flow is computed as the solution of a minimization problem. It is the first variational method for optical flow estimation and variational methods are still the predominant methods to estimate dense optical flow. From the assumption that pixel intensities do not change over time, the optical flow constraint equation is derived. This equation relates the optical flow with the derivatives of the image.

In practical situations, calculating dense optical flow is not simple: Let us consider for instance the motion of a white sheet of paper. In this case many of the white pixels in the previous frame will simply remain white in the next; only the edges may change, and even then only those perpendicular to the direction of motion. Hence the idea, developed originally in [22] of creating a sparse optical flow. This version of optical flow relies on some means of specifying beforehand the subset of points that are to be tracked. If these points have certain desirable properties, such as the 'corners,' then the tracking will be relatively robust and reliable.

In our acquisition environment, a hand may cross the whole acquired scene rather rapidly. This has lead us to consider adapting a pyramidal version [24] of Lucas-Kanade optical flow. This pyramidal version includes a multi-scale strategy which allows us to handle larger displacements while keeping the reduced computational load of Lucas-Kanade sparse method [22]. Moreover, in the past few years, optical flow has been widely improved to face harsh usage conditions, such as large displacements, by integrating rich descriptors [25] and to face discontinuities on motion boundaries [26].

In existing applications, optical flow is computed on the whole image in order to study its variations in brightness [21] or to provide a motion field [27].

The application which is the closest to our work is presented in [25]: The visual results exhibit video sequences of moving persons. For each sequence, a flow image is computed to determine the velocity of the features. However, as in previous works, the results obtained with optical flow are not interpreted in the sense that the movements of the persons in the scene are neither recognized nor labeled.

Unlike the conventional optical flow, in our work, we wish to restrict the computation of the optical flow to the fastest moving objects by rejecting the others. The novelty in our work relates to the development of a method to adapt optical flow to select a region of interest around an object independently of its size in order to facilitate the segmentation of the contour of the object. In practice, this region of interest will be delimited as the one

which surrounds the moving points detected by the optical flow. In our application the hand is the moving object in the scene. We expect that extracting a region of interest around the hand accelerates and facilitates the process of posture recognition, in terms of invariance to scaling for instance.

One difficulty of our application relates to the selection of a relatively small number of points detected by the optical flow. The norm of the corresponding flow vectors must fulfill the following conditions: it must be small enough to reject the unexpected vectors which are either due to variations in illumination; it must be high enough to reject the features which are slowly moving and which cannot be the hand. Another problem is how to get the 2-D gray level image required by optical flow from the RGB color image provided by the CMOS camera. In the general case, the mean value of the three R, G, and B channels is provided as an input image to the optical flow component. In specific conditions where one knows *a priori* the color of the hand, for example wearing a glove, it is easy to modify slightly this step of the algorithm and work on one of the R, G, or B channels.

3.3 Preprocessing steps for hand segmentation

Let N_{OF} be the number of moving points of interest, retrieved by the optical flow, from two frames: one obtained at time t, the other at time $t^{'} > t$. The coordinates of these points are denoted by $\{(x_o, y_o),\ o = 1, \ldots, N_{\mathrm{OF}}\}$.

Firstly, based on these points, we can select a region of interest (ROI) around the hand. As the shape of the hand is approximately elliptic, we choose an elliptic least-squares fitting [28]. The least-squares fitting methods are sensitive to outliers, and we wish to ensure the robustness of the ROI selection to variations in illumination. Therefore, we remove the moving points of interest which include an extreme (minimal or maximal) coordinate value. This may result in a possibly flawed detection by the optical flow. Let I^{p} denote an image containing the remaining moving points. Based on the fitting ellipse, we extract from the image a square ROI whose side length is 1.5 times the larger axis of the ellipse at time t.

Secondly, we segment the hand surface. We compute the center of mass of the moving points in I^{p}; then, we deduce the hand pixel gray level distribution in each RGB band from the region next to the center of mass. According to this distribution, we perform a histogram-based thresholding to each RGB band of the ROI as follows: for each band, the hand gray level distribution is computed around the center of mass. Let gl be the mean gray level value in this region. We compute a binary image where the pixels whose gray level values which are in the interval $[\mathrm{gl} - 5 : \mathrm{gl} + 5]$ are set to 1 while the others are set to 0. From these three binary images, we compute a 2-D binary image as the intersection of all 1-valued pixels of the thresholded R, G, and B bands. This binary image, denoted by I^{Th}, contains the hand surface filled with 1-valued pixels and noise, that is, 1-valued pixels randomly distributed in the image.

Thirdly, we isolate the hand surface and refine its shape: we remove the outliers and fill out the holes in I^{Th}. In practice, we select the largest set of connected pixels, assuming that this object is the hand. Then, we remove the unexpected background pixels which are connected to the hand with mathematical morphology operations, that is, erosions and dilations [1]. These operations remove the possibly remaining unexpected pixels from the background. This third preprocessing ensures that the whole algorithm is robust to variations in illumination and to the inclusion of unexpected objects in the background.

Fourthly, we select once again an ROI: the smallest square subimage containing the whole hand. The number of rows (columns) of this image are the vertical (horizontal) Ferret diameters [29] of the hand. We obtain an image denoted by I^{f} which contains only a filled hand.

Fifthly, we retrieve the hand contour, with a linear 'Roberts' filter. This yields an image I^{c} where the hand contour consists in 1-valued pixels, over a background of 0-valued pixels. This image will be used to compute a contour signature.

The preprocessing operations presented in this subsection allows us to focus on an ROI and isolate the hand contour. This ensures invariance properties of the signature which will be presented in the next section.

4 Characterization of the hand: a further investigation on a signature for non-star-shaped contours

A planar object shape can be characterized through two-dimensional moment invariants, obtained for instance with Hu [6], Zernike [7,30], or Legendre [31] moments. One-dimensional moment invariants can also be used as signatures to characterize contours, for instance Fourier descriptors [8,9], which are obtained by Fourier transform of the arclength parametrization, in complex coordinates, of a closed contour. An image scan in [32] provides a contour signature as a matrix involving the contour polar coordinates. An equivalent descriptor called shape context descriptor is presented in [33] as a compact human pose representation. The processed image is divided into different ranges of radial and angle coordinates. Each range tuple constitutes a bin. Counting the number of pixels in each bin yields a 2-D histogram. The main drawback of such a descriptor is that it does not provide a 1-pixel precision since it is impossible to distinguish between the pixels of a given bin; this may lead to details which are smaller than the bins

being skipped. Also, the more accurate the description, the smaller the regions, but the higher the computational load and the storage requirements. Here, we propose a contour signature which offers a resolution of 1 pixel.

The proposed scan is inspired not only from [32] but also from [34,35]. In [34] and [35], an image scan is proposed to characterize star-shaped contours. In a system of polar coordinates with adequately chosen poles, a contour is star-shaped if the radial coordinates (ρ) of its pixels are function of their angular coordinates (θ): for one θ value there is only one ρ value $\rho = f(\theta)$. In the general case, hand contours are not star-shaped since it is (nearly) impossible to find a pole for which the relation $\rho = f(\theta)$ holds for all contour pixels. This has inspired us to investigate a characterization method which handles non-star-shaped contours.

The proposed method for contour characterization splits the image into several rings centered on a reference point. The requirements on the location of this reference point are low, unlike the condition imposed by the method in [35]. With this characterization method, we aim at distinguishing very similar postures with a computational load which is lower than what the generally used Fourier descriptors would require.

An image I^c, denoted by I, is assumed to have a size $N \times N$, and its pixels are referred to, starting from the top left corner of the image, as $I_{l,m}$ (see Figure 3). The 1-valued pixels constitute the expected contour where the contour pixels are located in a system of polar coordinates with pole $\{l_c, m_c\}$ (see Figure 3). Unlike the methods proposed in [35], where the center must be chosen in such a way that the contour is star-shaped, the computation of the center coordinates is not essential. For instance, this pole can be the center of mass obtained in the previous section. What we call signature in this paper is a set of data which characterize the corresponding contour. The signature that we investigate in this paper is based on the generation of signals out of an image. As in [35], a circular array of sensors is associated with the image. The sensor array is supposed to be placed along a circle centered on the pole $\{l_c, m_c\}$. The number of sensors is denoted by Q and one sensor corresponds to one direction for the signal generation D_i, which makes an angle θ_i with the vertical axis. See for instance the ith and the Qth sensors in Figure 4. The other sensors are not represented for the sake of clarity.

The method proposed in [35] is valid only for contours exhibiting at most 1 pixel for one direction D_i. We wish to overcome this limitation and characterize non-star-shaped contours, because the hand contours considered in this paper are mostly non-star-shaped. To separate the influence of each pixel located along a given direction D_i, we no longer generate one 1D signal, but a number P of 1D signals on the antenna. Each signal corresponds to one 'ring' represented on Figure 4. We assume that for each direction D_i, there is only 1 pixel in each of the P intervals. Because I is square and of course not circular, P differs

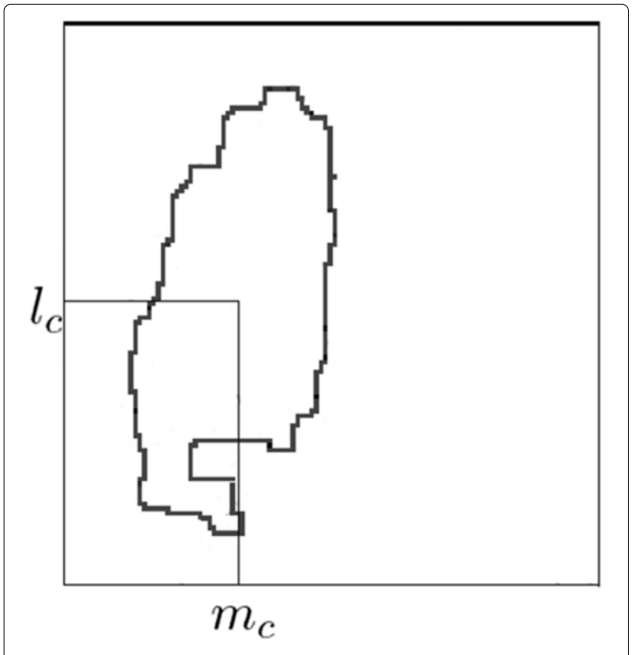

Figure 3 Image and edge model.

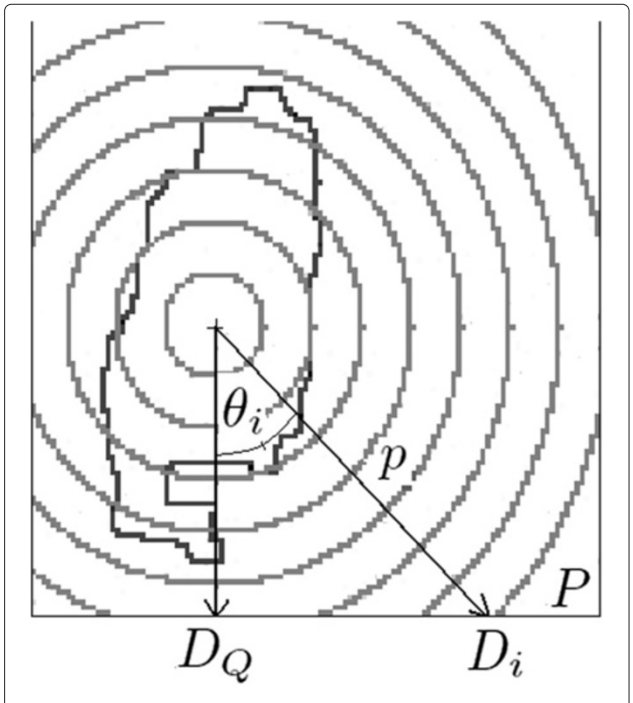

Figure 4 Signal generation process.

from one direction, D_i to another. Its maximum theoretical value is, for instance, $\frac{N}{\sqrt{2}}$, if $l_c = N/2$ and $m_c = N/2$. In these conditions, the value of Q should not exceed $\sqrt{2}\pi N$ and it is sufficient to take into account all pixels of a given interval p. Therefore, we generate P signal vectors for each direction D_i. For the pth interval ($p = 1, \ldots, P$) and the direction D_i ($i = 1, \ldots, Q$), the signal component $z_{p,i}$ is computed as follows:

$$z_{p,i} = I_{l_{p,i}, m_{p,i}} \sqrt{(l_{p,i} - l_c)^2 + (m_{p,i} - m_c)^2} \quad (1)$$

In Equation 1, $l_{p,i}$ and $m_{p,i}$ are the coordinates of the pixel located in the pth interval and along the ith direction D_i; $I_{l_{p,i}, m_{p,i}}$ is the pixel value, either 0 or 1, of I at the location $l_{p,i}, m_{p,i}$, and we remind that $\{l_c, m_c\}$ are the coordinates of the pole from which the directions D_i for signal generation start. The components $z_{p,i}$ of the signature are equal to the distance to the pole of the 1-valued pixels.

These components $z_{p,i}$ ($p = 1, \ldots, P$, $i = 1, \ldots, Q$) can be grouped into a matrix $\mathbf{Z}$ of size $P \times Q$:

$$\mathbf{Z} = \begin{bmatrix} z_{1,1} & z_{1,2} & \cdot & \cdot & z_{1,Q} \\ z_{2,1} & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & z_{p,i} & \cdot & \cdot \\ z_{P,1} & \cdot & \cdot & \cdot & z_{P,Q} \end{bmatrix} \quad (2)$$

The first rows of matrix $\mathbf{Z}$ correspond to the first intervals which are the nearest to the pole (see Figure 4). The values of the non-zero components $z_{p,i}$ in these rows will then be rather small. Conversely, the last rows of matrix $\mathbf{Z}$ will contain large values. One can notice also that for closed contours, any component $z_{p,Q}$ of the last column differs from the component of the same row and the first column $z_{p,1}$ by at most 1 pixel.

All columns of $\mathbf{Z}$ should have the same number of rows so that for the directions D_i which cross less than P intervals, 0-valued components are set in $\mathbf{Z}$ for the corresponding indices i. If the width of the intervals is chosen such that there is at most 1 pixel per direction D_i and per interval, this matrix allows us to reconstruct exactly the contour; it contains the radial coordinates of the contour in the system of pole $\{l_c, m_c\}$. However, the purpose of the signature is not necessarily to reconstruct exactly the contour; it should characterize a contour so that all postures can be distinguished. Also, the signature should be invariant to rotation. To achieve this, in previous works [3,4], we proposed to straighten up the image through several rotations, in order to maximize the hand Feret's diameter in the horizontal direction. Instead, in this paper, we propose a technique which is much less computationally intensive where the components $z_{p,i}$ of a given interval p are sorted. Consequently, all non-zero values of the pth row of $\mathbf{Z}$, where each corresponds to a contour pixel, are ordered and turned as the last components of the pth row.

The proposed signature has invariant properties which are summarized follows:

- It is invariant to translation: for any position of the hand in the initial image, the box which encloses the contour is blindly estimated.
- It is invariant to scaling: whatever the size of the region of interest (small number of pixels if the camera is far from the hand, large number of pixels if the camera is near to the hand), the number of intervals P for the radial coordinate values, and the number Q of directions for signature generation is always the same. As a consequence, the size of matrix $\mathbf{Z}$ will be constant, whether the user's hand is

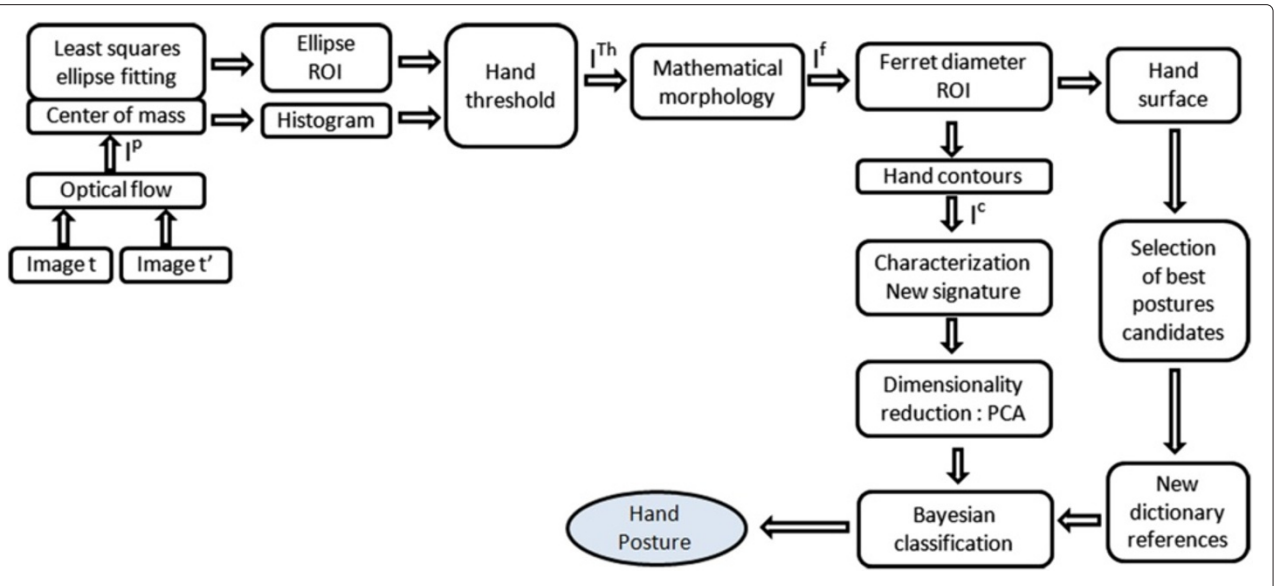

Figure 5 Proposed algorithm for hand posture recognition.

near to or far from the camera. The signature depends on the shape of the hand, not on its size.

• It is invariant to rotation: for a given image, applying a rotation leaves the number of non-zero pixels in the pth interval and their distances to the pole $\{l_c, m_c\}$ unchanged. Only the column index of the components $z_{p,i}$ in the signature is modified after rotation. After sorting, the components $z_{p,i}$ for both images, before and after rotation, are stored in the same columns of the signature. So, the sorting described above ensures the invariance to rotation.

• Sorting the components of each row p of matrix $\mathbf{Z}$ permits to handle left hands as well as right hands. For a given posture, the user is allowed to use his left hand or his right hand, because the only difference observed in the signature before sorting between the two cases resides in the column indices of the components $z_{p,i}$ for a given row p. After sorting, the signature $\mathbf{Z}$ is the same for a right and for a left hand. This is an important progress with respect to previous works [3,4,10] because it relaxes the conditions of use of our method.

These invariance properties allow us to employ the proposed contour signature for hand posture classification purposes.

5 Classification of the hand posture

Let us consider H classes of hand postures. For the purpose of hand posture classification, Euclidean and Mahalanobis distances are used in [3,4]. In this paper we will focus on Mahalanobis distance and use the Euclidean distance as comparative method. We vectorize a matrix $\mathbf{Z}$ characterizing a posture into a $P.Q$ vector denoted by $\mathbf{x}$. For each class h, a subset of hand photographs is available. The H subsets constitute the learning set. This set was created by an expert who knows exactly what position his fingers should have to fit each posture in Figure 1. For a given class h, let M_h be the number of images for this class in the learning set. Let $\mathbf{x}_{n_h}$, $n_h = 1, \ldots, M_h$ be the vectors '$\mathbf{x}$' obtained from the images belonging to class h after computing the matrix $\mathbf{Z}$. Let $\mathbf{X}_h$ be the matrix whose columns are the vectors $\mathbf{x}_{n_h}$, $n_h = 1, \ldots, M_h$. It is obvious from Figure 4 that the higher P and Q, the more details are captured by the signature $\mathbf{Z}$, the more accurate the hand posture classification method involving this signature is. However, for large values of P and Q, $\mathbf{X}_h$ exhibits a large number of rows, and it is a sparse matrix. For a given class h, the matrix $\mathbf{X}_h$ allows us to extract some 'general' or mean characteristics of the posture class h. The principles of posture classification are as follows: a test set is created from persons who are not the expert, who may wear

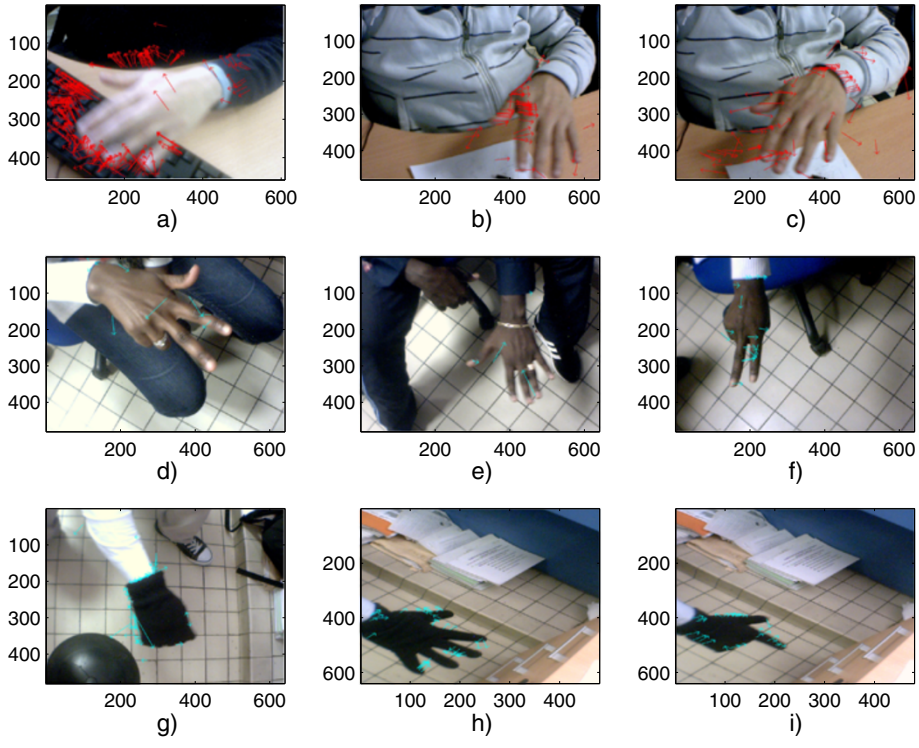

Figure 6 (a,b,c) White-skinned people. (d,e,f) Colored people. **(g,h,i)** Black gloves. Motion detection with optical flow. The red and light blue arrows inform about the hand movement: their length is proportional to speed and their direction is the same as the movement's one.

color gloves, or be dark-skinned. We aim at associating a label with any image chosen from the test set. This label is one of the 11 postures presented in Section 2. To improve the recognition rate with respect to the work presented in [3,4], we propose to reduce the number of candidates for a posture with a sphericity criterion. In Subsection 5.1, we reduce the dimensionality of matrix $\mathbf{X}_h$ obtained from the learning set.

5.1 Dimensionality reduction and Mahalanobis distance computation

As explained previously, the matrices $\mathbf{X}_h$ obtained from the learning set may have a large number of rows and as such may be sparse. Therefore, a dimensionality reduction of the data is useful. The projection pursuit (PP) [36] can be used since it is an iterative algorithm which can be faster than the more conventional principal component analysis (PCA) which is based on singular value decomposition. However, the model provided by PP is often difficult to interpret because a PP tends to yield any distribution for the representative data but the normal one. On the other hand, PCA represents the data as a combination of vectors along which the data exhibited by the highest variability (or variance). PCA expresses the data as a linear combination of the most significant features of a given posture. Also, a fast and iterative version of PCA is also convenient. It uses fixed point algorithm instead of singular value decomposition [37]. For these useful reasons, we have chosen a fast version of PCA to perform dimensionality reduction.

Let K ($K < P.Q$) be the number of dominant singular values in $\mathbf{X}_h$. Let $\mathbf{U}_h$ be the matrix whose columns are the K singular vectors associated with the K largest singular values of $\mathbf{X}_h$. The singular vectors stored in $\mathbf{U}_h$ are computed with fixed point algorithm [38], which is much faster than the classical singular value decomposition when $K \ll P.Q$ (which is the case in the considered paper). Each singular vector corresponds to a mode of variation of the considered hand posture of class h, and its corresponding singular value is related to the variance along the mode specified by the singular vector. The compressed version of the data is obtained by $\mathbf{X}_h^c = \mathbf{U}_h^{\mathrm{T}}\mathbf{X}_h$, where $^{\mathrm{T}}$ denotes transpose. Let $\mathbf{x}_{n_h}^c$, $n_h = 1,\ldots,M_h$ denote the columns of $\mathbf{X}_h^c$. The columns of matrix $\mathbf{X}_h$ characterize the hand postures in the corresponding images independently from their location and size, owing to the invariance of the proposed signature $\mathbf{Z}$ to translation, rotation, and scaling. So then are the columns of matrix $\mathbf{X}_h^c$. We compute the mean and the covariance of the columns of $\mathbf{X}_h^c$ to obtain the reference characteristics for any posture h. The mean vector is computed as

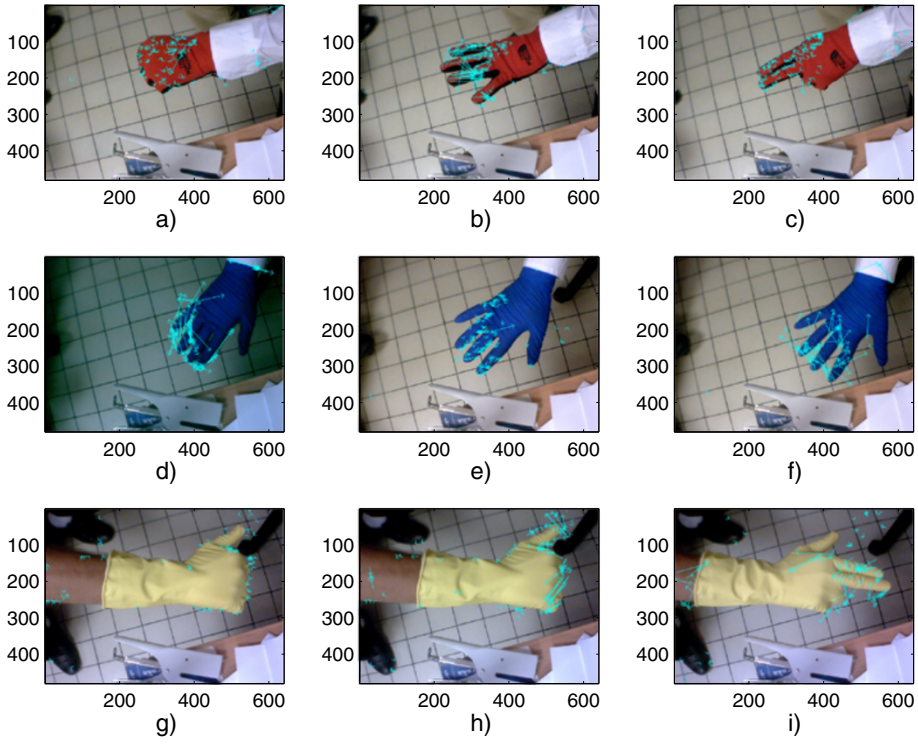

Figure 7 Motion detection with optical flow. (a,b,c) Red hand. **(d,e,f)** Blue hand. **(g,h,i)** Yellow hand. The light blue arrows inform about the hand movement: their length is proportional to speed and their direction is the same as the movement's one.

$\boldsymbol{\mu}_h = \frac{1}{M_h}\sum_{n_h=1}^{M_h}\mathbf{x}_{n_h}^c$, and the covariance matrix is computed as $\boldsymbol{\Lambda}_h = \frac{1}{M_h}\sum_{n_h=1}^{M_h}(\mathbf{x}_{n_h}^c - \boldsymbol{\mu}_h)(\mathbf{x}_{n_h}^c - \boldsymbol{\mu}_h)^{\mathrm{T}}$, for each class $h = 1,\ldots,H$. Equivalently, the covariance matrix of the reduced data can be expressed as follows: $\boldsymbol{\Lambda}_h = (\mathbf{U}_h^{\mathrm{T}}\mathbf{X}_h - \boldsymbol{\mu}_h)(\mathbf{U}_h^{\mathrm{T}}\mathbf{X}_h - \boldsymbol{\mu}_h)^{\mathrm{T}}$. Even if there are small variations from one posture provided by the expert to another, these variations are smoothed through the computation of the mean invariant vector $\boldsymbol{\mu}_h$. Any image coming from the test set and characterized by vector $\mathbf{x}$ is classified by minimizing a distance. Two main distances may be chosen: the Euclidean distance or the Mahalanobis distance which is applied to the compressed vector $\mathbf{U}_h^{\mathrm{T}}\mathbf{x}$ as follows:

$$\mathcal{D}_m = (\mathbf{U}_h^{\mathrm{T}}\mathbf{x} - \boldsymbol{\mu}_h)^{\mathrm{T}}(\boldsymbol{\Lambda}_h + \epsilon)^{-1}(\mathbf{U}_h^{\mathrm{T}}\mathbf{x} - \boldsymbol{\mu}_h) \tag{3}$$

where ϵ is a small-valued regularization constant, fixed by the user and added to prevent any numerical instability while inverting the matrix $\boldsymbol{\Lambda}_h$. Computing the Mahalanobis distance involves, as shown in Equation 3, the inversion of the covariance matrix $\boldsymbol{\Lambda}_h$. This is not the case for the Euclidean distance which is then easier to implement than the Mahalanobis distance, but the Mahalanobis distance usually provides better classification results; this has been verified in the frame of hand posture recognition in [13]. This comes from the influence of the factor $\boldsymbol{\Lambda}_h$ which scales the influence of each feature characterizing any posture h. Consequently, we propose to use the Mahalanobis distance. From the definition above, matrix $\boldsymbol{\Lambda}_h$ is of dimension $K \times K$. With the compressed version of the data, we have obtained a lower-dimensional representation of the reference hand posture h, which is more suitable to describe any test posture. Each dimension describes a natural mode of variation of how the user presents its hand in posture h in front of the camera.

In addition, the dimensionality reduction has lead to a reduction of the computational load dedicated to matrix inversion in Equation 3 where the matrix $\boldsymbol{\Lambda}_h$ was computed from the compressed data and has low $K \times K$ size if the chosen K is small enough. This also prevents from inverting an ill-conditioned matrix. For the sake of comparison, the proposed signature can be also exploited with a Euclidian distance which can be computed as follows: $||\mathbf{U}_h^{\mathrm{T}}\mathbf{x} - \boldsymbol{\mu}_h||$, where $||.||$ denotes Frobenius norm.

5.2 Preselection of best posture candidates

Through a careful analysis of the dictionary of posture (see Figure 1), we can distinguish two large categories of

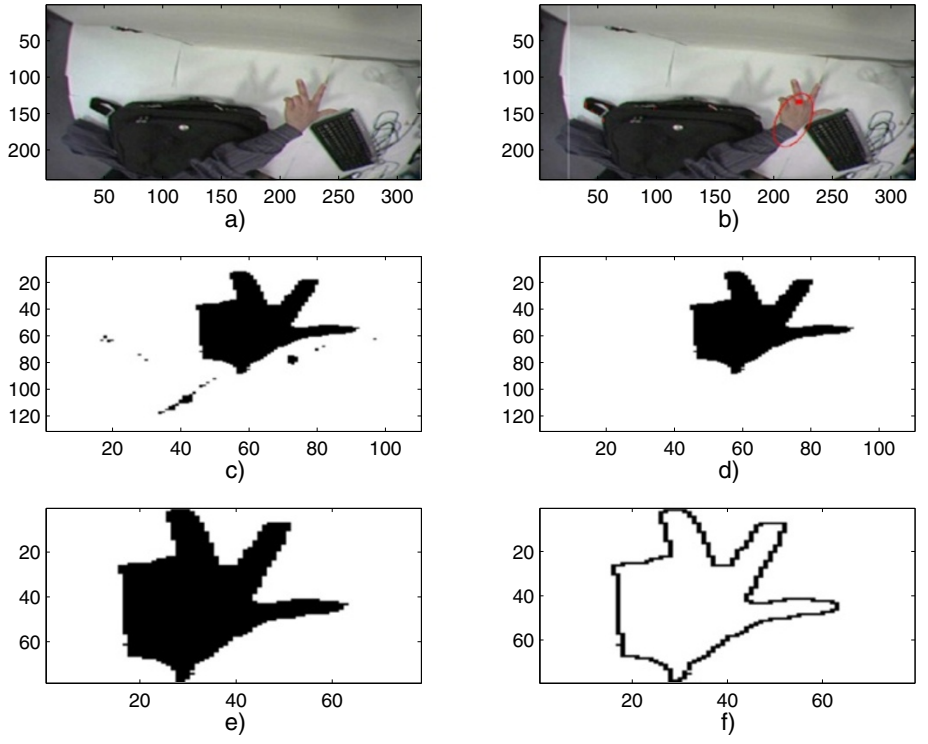

Figure 8 White hand and steps of the proposed method. (a) Processed image. **(b)** Moving points, fitting ellipse (indicated by red dots and a red ellipse), and center of mass (indicated by a red square). **(c)** Threshold image I^{Th} in the ROI defined from the fitting ellipse. **(d)** Result obtained after mathematical morphology operations. **(e)** I^{f}-square ROI whose height is the maximum Ferret diameter of the hand. **(f)** I^{c}, obtained from Roberts linear filtering, containing the expected hand contour.

postures. To characterize them, we introduce a sphericity criterion, denoted by S, which is low when the posture contour exhibits concavities and grows, tending to 1, when the posture contour tends to be a circle. The computation of S involves the hand surface computed from I^{f} and the length of the hand contour, computed from I^{c}: $S = \frac{4\pi\mathcal{N}(I^{\mathrm{f}})}{(\mathcal{N}(I^{\mathrm{c}}))^2}$, where $\mathcal{N}(\cdot)$ denotes the number of 1-valued pixels. For instance, postures 1, 4, 5, 6, and 10 exhibit a rather high sphericity criterion while postures 2, 3, 8, 9, and 11 a rather low sphericity criterion. Posture 7 lies in between the two. The advantage of the sphericity criterion with respect to the surface criterion proposed in [4] is that it guarantees the invariance to translation and also to scaling and rotation. Our aim is then to preselect a group of six postures, rejecting the other five postures to which a considered test posture does not belong to.

This is done in the following way: the criterion S is computed for all images of each class h in the learning set. Let S_h be the mean value of the sphericity criterion for a given class h. Let us consider a test image with a sphericity criterion S_t. We wish to select the six postures which best represent the test image in terms of sphericity criterion. The criterion S is scalar, real-valued and positive, so the l_1 norm is the most adequate to compute the distance between the sphericity criterion of the test image and the mean sphericity criterion for all classes: namely, the distance $|S_t - S_h|$ between S_t and S_h is computed for all classes h, and we select the six classes which yield the minimum criterion value. This allows us to build a new dictionary with a reduced number of candidates, and the distance $\mathcal{D}_m$ of Equation 3 is computed only six times to perform classification.

6 Hand posture recognition: summary and numerical complexity of the proposed methods

6.1 Summary of the proposed methods

Figure 5 depicts the overall structure of our algorithm. Referring to an analogous scheme in [4], one can appreciate some improvements that were recently carried out to further enhance the detection of hand contours, essentially the optical flow combined with elliptic fitting, a dedicated preprocessing to help in the selection of the largest set of connected pixels, and a new method to ensure the invariance to rotation.

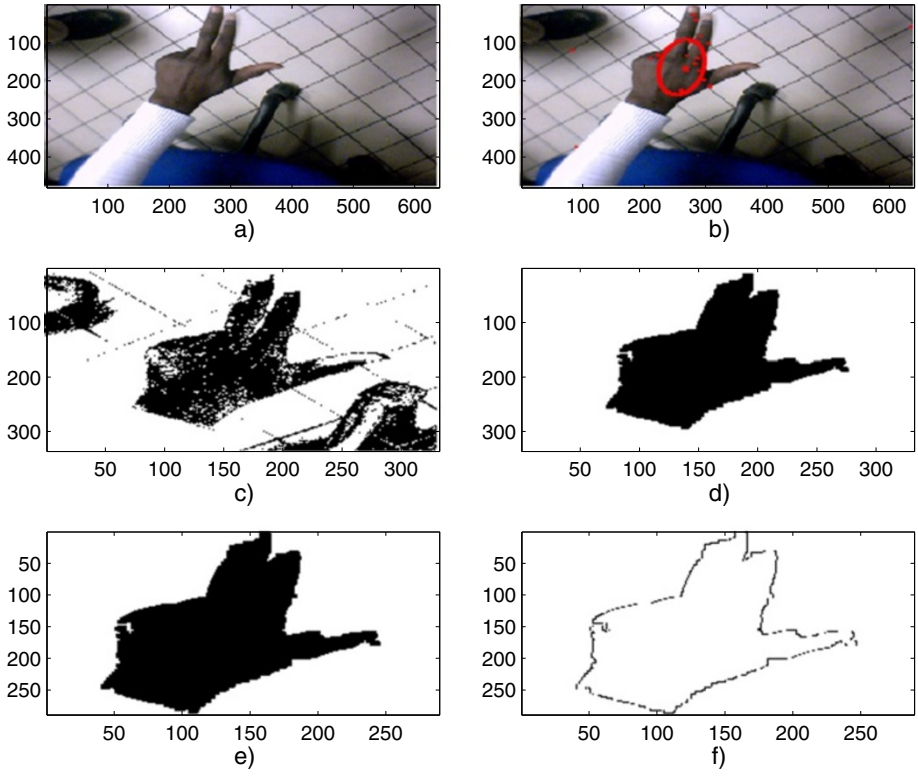

Figure 9 Hand of a colored person and steps of the proposed method. (a) Processed image. **(b)** Moving points, fitting ellipse (indicated by red dots and a red ellipse), and center of mass (indicated by a red square). **(c)** Threshold image I^{Th} in the ROI defined from the fitting ellipse. **(d)** Result obtained after mathematical morphology operations. **(e)** I^{f}-square ROI whose height is the maximum Ferret diameter of the hand. **(f)** I^{c}, obtained from Roberts linear filtering, containing the expected hand contour.

6.2 Numerical complexity of the proposed methods

To express the computational complexity of the proposed method, let us assume that if f denotes the actual computational complexity depending on the parameters of the methods, and if g is another function of these parameters, we can use the notation $f = O(g)$, if and only if there exist a positive constant a such that for all sufficiently large values of the parameters, the numerical complexity f is at most a multiplied by g in absolute value. Therefore, we will use this notation by considering the case where the computational load is the highest.

We will consider successively the methods applied both off-line and on-line, only off-line, and then only on-line.

• The following steps are applied both off-line to build the reference database, and on-line to identify a hand posture. The first one, optical flow, is based on the computation of subsampled images from the two images available at time t and at time $t + 1$. Let N_t be the largest dimension of the initial image (for instance the number of rows for a tall image) at time t or at time $t + 1$. Spatial derivatives are computed for each subsampled image. These are pixel-to-pixel operations based on subtractions, so the order of magnitude of such computations is $O({N_t}^2)$ operations (at most a multiple of the number of pixels). The subsequent preprocessing operations are performed on a ROI extracted from image t. The largest dimension of this ROI is marginally larger than, or equal to N, so at most a multiple of N (see Subsection 3.3). These preprocessing operations are the least-squares ellipse fitting which is not computationally dominant, and then the mathematical morphology filters, yielding a computational complexity of $O(sN^2)$, where s is the number of pixels in the structuring elements [39]. As s is fixed for any image, this complexity is equivalent to $O(N^2)$. To compute the proposed signature, computing and storing one element of signature $\mathbf{Z}$ requires three multiplications, two subtractions, one addition, and one square root. Assuming that multiplications and square root are the most computationally intensive/costly and assuming that in the worst case, 1 pixel is present for each level p and for each direction of generation D_i, this yields a complexity of $4\ P\ Q$, that is, an order of magnitude of $O(PQ)$. Sorting one row of matrix $\mathbf{Z}$ requires $O(Q\log(Q))$ operations, and sorting all rows requires $O(PQ\log(Q))$ operations.

• The following steps are only applied off-line: the computation of the mean characteristics $\mathbf{U}_h$, $\boldsymbol{\mu}_h$ for each of

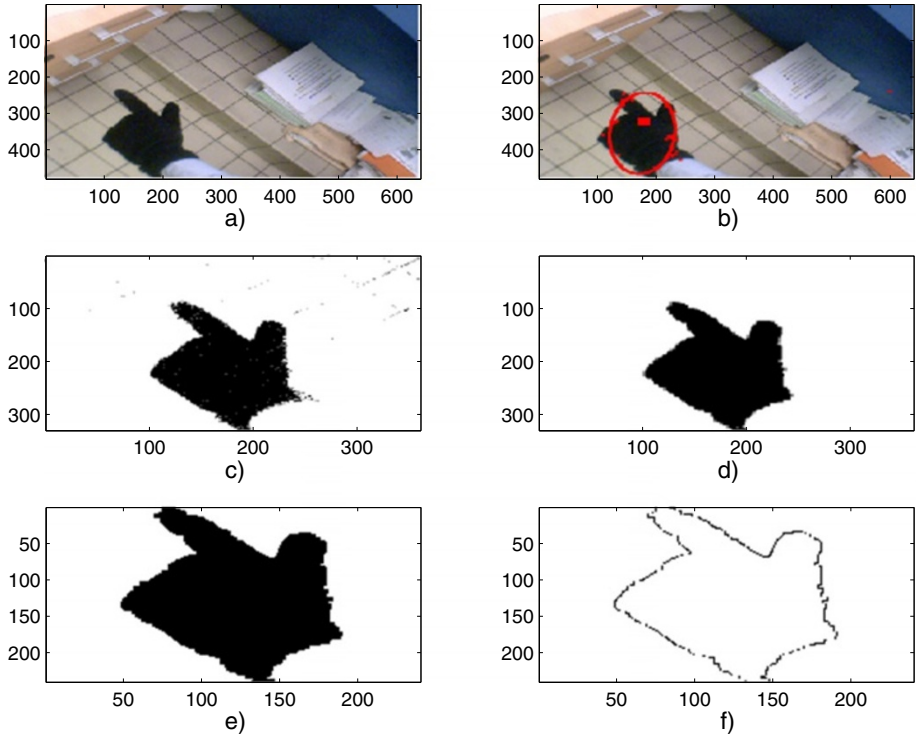

Figure 10 Hand wearing a black glove and steps of the proposed method. (a) Processed image. **(b)** Moving points, fitting ellipse (indicated by red dots and a red ellipse), and center of mass (indicated by a red square). **(c)** Threshold image I^{Th} in the ROI defined from the fitting ellipse. **(d)** Result obtained after mathematical morphology operations. **(e)** I^{f}-square ROI whose height is the maximum Ferret diameter of the hand. **(f)** I^{c}, obtained from Roberts linear filtering, containing the expected hand contour.

the 11 hand postures h is performed by PCA, which relies on the computation of leading singular vectors. For this purpose we use the fixed point algorithm (see Subsection 5.1). This algorithm yields a complexity $O(K(PQ)^2 + M_h K(PQ)^2)$ [37], where M_h is the number of sample images for the considered class, and K the number of leading singular vectors. The inversion of matrix $\mathbf{\Lambda}_h$ by Gauss Jordan elimination requires $O(K^3)$ operations.

• The following steps are only applied on-line: for any test image, as there exist 11 candidate hand postures h, the computation of the 11 distances between the scalar sphericity criterion S_t of the test image and the 11 mean sphericity criteria S_h. After preselection of six postures, the six Mahalanobis distance values $\mathcal{D}_m$ between the compressed test vector $\mathbf{U}_h^{\mathrm{T}}\mathbf{x}$ and the mean vector $\boldsymbol{\mu}_h$. This yields $11 + 6O(K)$ operations. Additionally, a sorting operation is performed for each distance, with an associated computational complexity $11\log(11)$ and $6\log(6)$.

Finally, $O({N_t}^2 + N^2 + PQ + PQ\log(Q) + 11(K(PQ)^2 + M_h K(PQ)^2 + K^3))$ operations are performed off-line, and $O({N_t}^2 + N^2 + PQ + PQ\log(Q) + 6K) + \delta$ operations, where δ is a constant not depending on the parameter values, which are performed on-line. Considering sufficiently large values of N_t, N, P, Q, and M_h, we derive the following simplified expressions:

- $O({N_t}^2 + N^2 + PQ(\log(Q) + KM_h PQ) + K^3)$ operations off-line;
- $O({N_t}^2 + N^2 + PQ\log(Q) + K)$ operations on-line.

Eventually, we can further simplify these expressions: in the processed image, the ROI containing the hand is large enough to ensure $N > \frac{N_t}{10}$. So, $O(N^2)$ is equivalent to $O({N_t}^2)$, and we can derive the following simplified expressions:

- $O(N^2 + PQ(\log(Q) + KM_h PQ) + K^3)$ operations off-line;
- $O(N^2 + PQ\log(Q) + K)$ operations on-line.

Firstly, we can notice that the computational complexity of the off-line methods is of one order of magnitude higher in PQ. This is due to the computation of the mean characteristics $\mathbf{U}_h$, $\boldsymbol{\mu}_h$, which is only performed off-line. The

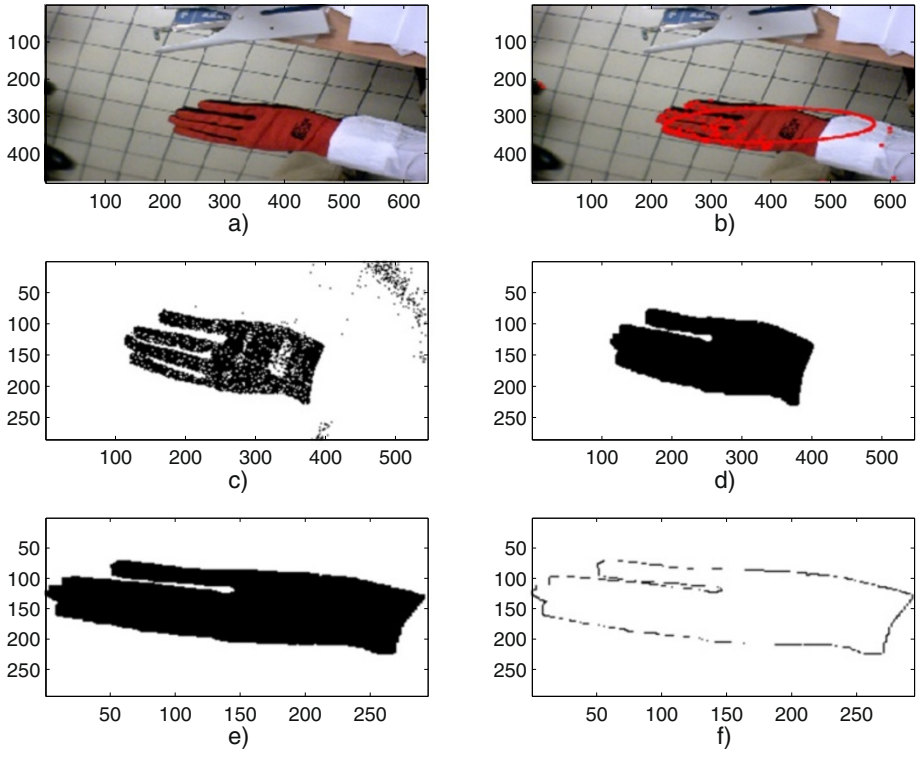

Figure 11 Red hand and steps of the proposed method. (a) Processed image. **(b)** Moving points, fitting ellipse (indicated by red dots and a red ellipse), and center of mass (indicated by a red square). **(c)** Threshold image I^{Th} in the ROI defined from the fitting ellipse. **(d)** Result obtained after mathematical morphology operations. **(e)** I^{f}-square ROI whose height is the maximum Ferret diameter of the hand. **(f)** I^{c}, obtained from Roberts linear filtering, containing the expected hand contour.

off-line operations also include the term K^3, because the inverse of each covariance matrix is performed off-line. These computations have no influence on the speed of the hand posture recognition process. Secondly, we remind that the computational load required in practice for each step depends on the programming language which was used to implement it.

In Section 7, the results obtained are presented, in terms of recognition and also computational time performance.

7 Results on hand posture recognition and discussion

In the database, the images have a size of either 320 × 240 or 640 × 480 depending on the camera used at the acquisition step. The machine used has a 2-core processor running at 3.2 GHz, using Matlab®. This result section falls into two subsections: firstly, we present the results of hand contour segmentation with optical flow and adequate preprocessing; secondly, we exemplify the discriminative capabilities of the signature **Z** and present the statistical results of hand posture recognition obtained with PCA and Mahalanobis distance.

7.1 Hand contour segmentation with optical flow

We present successively some requirements for the optical flow to work optimally: an initial exemplification of optical flow method applied to a white and a dark-skinned hand and to a colored moving hand wearing a glove and a second exemplification of all preprocessing methods, including optical flow, which yield the binary image I containing the hand contour. This binary image is smaller than the input color images, and its size depends on the distance between the hand and the camera. We have chosen three levels for the pyramids while applying optical flow.

7.1.1 Requirements for the experimental conditions and subsequent image rejection

Adapting optical flow exhibits not only advantages but also requirements on the experimental conditions and specific preprocessing. The required experimental conditions for which the optical flow works optimally are as follows [22]: between two frames, the hand should move slowly enough to avoid blurred images and the brightness must be constant, so that a moving pixel does not change in appearance; to avoid the presence of odd

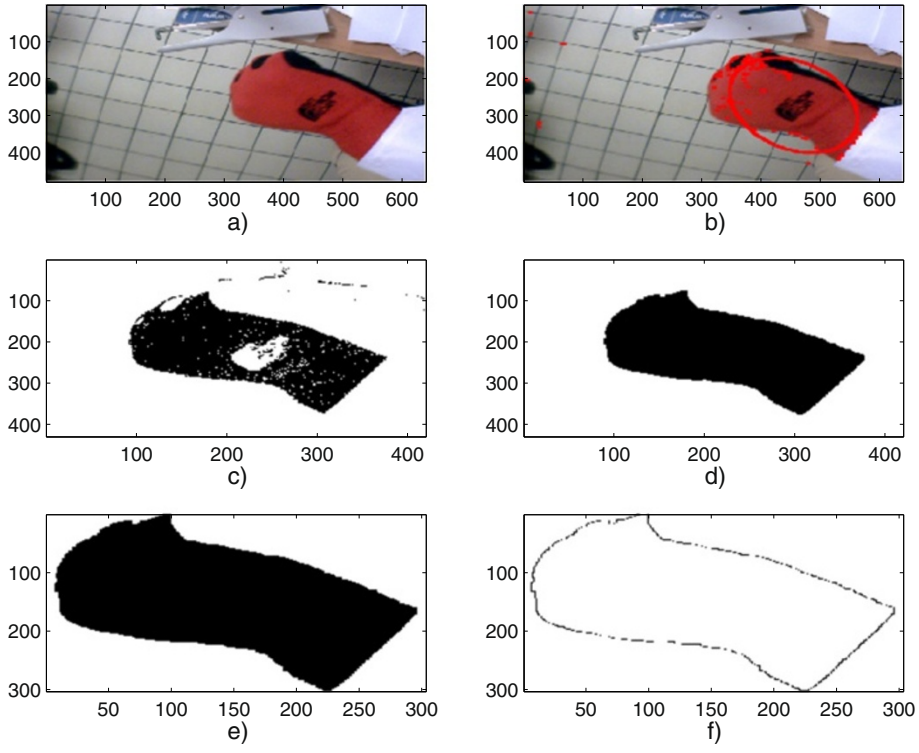

Figure 12 Red hand 2 and steps of the proposed method. (a) Processed image. **(b)** Moving points, fitting ellipse (indicated by red dots and a red ellipse), and center of mass (indicated by a red square). **(c)** Threshold image I^{Th} in the ROI defined from the fitting ellipse. **(d)** Result obtained after mathematical morphology operations. **(e)** I^f-square ROI whose height is the maximum Ferret diameter of the hand. **(f)** I^C, obtained from Roberts linear filtering, containing the expected hand contour.

points among the set of moving points detected by optical flow, the background color must be different from the hand color, which is easy to set including with complex backgrounds.

We comply with these requirements and cope with odd detected points as follows:

- Firstly, while running optical flow, we choose a time increment between two frames which is small enough. To reduce the computational load and reach a delay between two subsequent frames which is as low as possible, we set the best empirically chosen parameters: we reduced the number of moving points taken into account by optical flow to 1,000, which permitted to reach a time delay of 1 10^{-1} s between two frames.
- Secondly, while creating our learning database, we reject the frames which yield such characteristics: if one axis of the fitting ellipse (see Subsection 3.3) is larger than the image size, or if the large axis is larger than three times the small axis. We consider that in these cases, the moving points detected by optical flow do not permit to delimitate a reliable region of interest around the hand.

The drawback/limitation for a user point of view is that the hand posture recognition may take a bit longer for the recognition result, until the luminosity does not vary too much, or until his hand moves slowly enough.

7.1.2 Performance assessment on colored hands

The main advantages of the proposed method, which adapts optical flow instead of the classically used YC_bC_r mapping, are as follows: it handles the case of colored hands, such as people wearing gloves of any color, or hands of dark-skinned people. This is a significant advantage over the existing methods which usually fail as soon as the hand surface cannot be distinguished from the background in the C_r component. See for instance the issues concerning the background presented in [11,18-20] and the problem of body parts which are present in the scene and which are not the hand [14]. Also, we adapt optical flow to isolate a region of interest containing the hand, which reduces the computational load of the subsequent steps. Conversely, in comparative methods, when YC_bC_r mapping is performed, the threshold operation is applied to the whole image, which is computationally more costly, and as such increases the probability to get noise pixels.

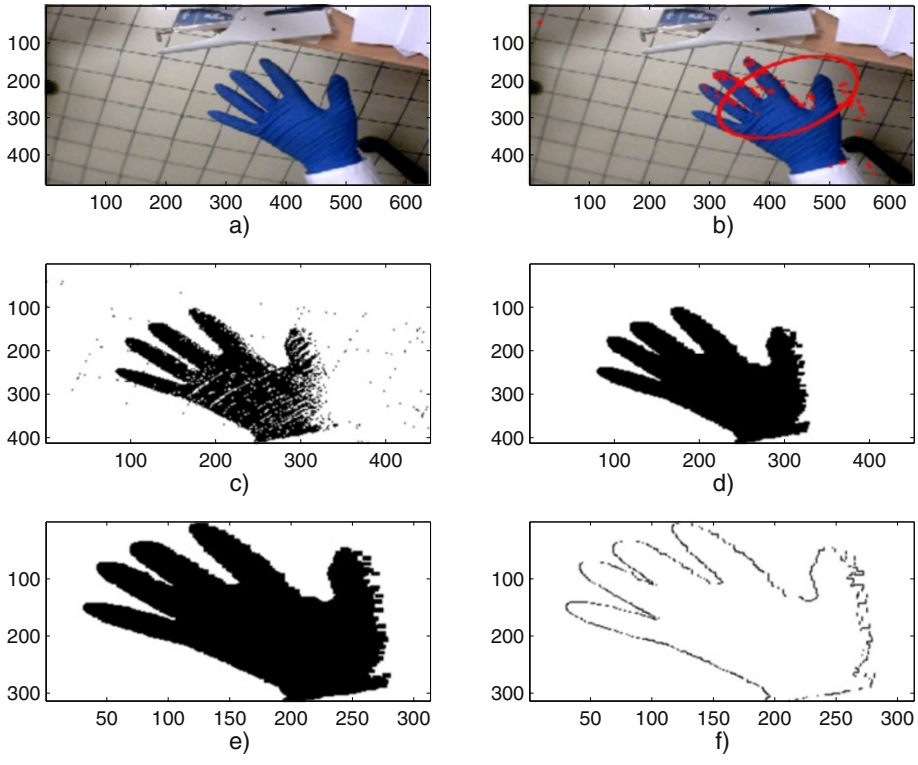

Figure 13 Blue hand and steps of the proposed method. (a) Processed image. **(b)** Moving points, fitting ellipse (indicated by red dots and a red ellipse), and center of mass (indicated by a red square). **(c)** Threshold image I^{Th} in the ROI defined from the fitting ellipse. **(d)** Result obtained after mathematical morphology operations. **(e)** I^{f}-square ROI whose height is the maximum Ferret diameter of the hand. **(f)** I^{c}, obtained from Roberts linear filtering, containing the expected hand contour.

Figure 6 shows the results obtained by the optical flow on hands from white people (Figure 6a,b,c), dark-skinned people (Figure 6d,e,f), and hands wearing black gloves (Figure 6g,h,i); Figure 7 shows the results obtained by optical flow on a red (Figure 7a,b,c), a blue (Figure 7d,e,f) and a yellow hand (Figure 7g,h,i). It consists of pixels which are about to move between the current and the next frame. It can be expected from these pixels that they delimitate correctly an ROI around the moving hand: the results obtained by the optical flow can be used to perform the preprocessing step.

As shown in Figures 6 and 7, the optical flow method provides a set of points, among the moving points of the scene. Since a sparse version of the optical flow was chosen, these points are mainly focused on the hand contour. In Figure 8 we exemplify the steps of the proposed algorithm for the segmentation of hand contour, on a white-skinned hand, and a posture of type '3'. This algorithm starts with an adaptation of the optical flow and includes a thresholding operation. In Figure 8a, we show the processed image. The five other figure parts in Figure 8 illustrate the five preprocessing steps of Subsection 3.3. In Figure 8b, we show the moving points provided by the optical flow (colored in red), their center of mass (as a large red square), and the fitting ellipse (also in red). The image I^{Th} of Figure 8c is the intersection of all 1-valued pixels of the thresholded R, G, and B bands. This thresholding operation is selective enough to isolate correctly the hand surface, but noise pixels appear in I^{Th} as illustrated in Figure 8c. Figure 8d illustrates the advantages of the third preprocessing presented in Subsection 3.3: it allows the removal of the undesired pixels which are present in the thresholded image I^{Th}. We present in Figure 8e the ROI which is the smallest square image containing the hand. In Figure 8f, we present the hand contour obtained by Roberts linear filtering applied to the image of Figure 8e.

In Figures 9, 10, 11, 12, 13, 14, 15, 16, we further exemplify the method in the same way, with a dark-skinned hand, and a hand wearing a black, a red, a blue and a yellow glove.

The results obtained on these sample images show the ability of the proposed method to handle not only white but also colored hands.

7.2 Hand posture characterization and recognition

Some experiments performed with our database allowed us to determine and set the best values for parameters P

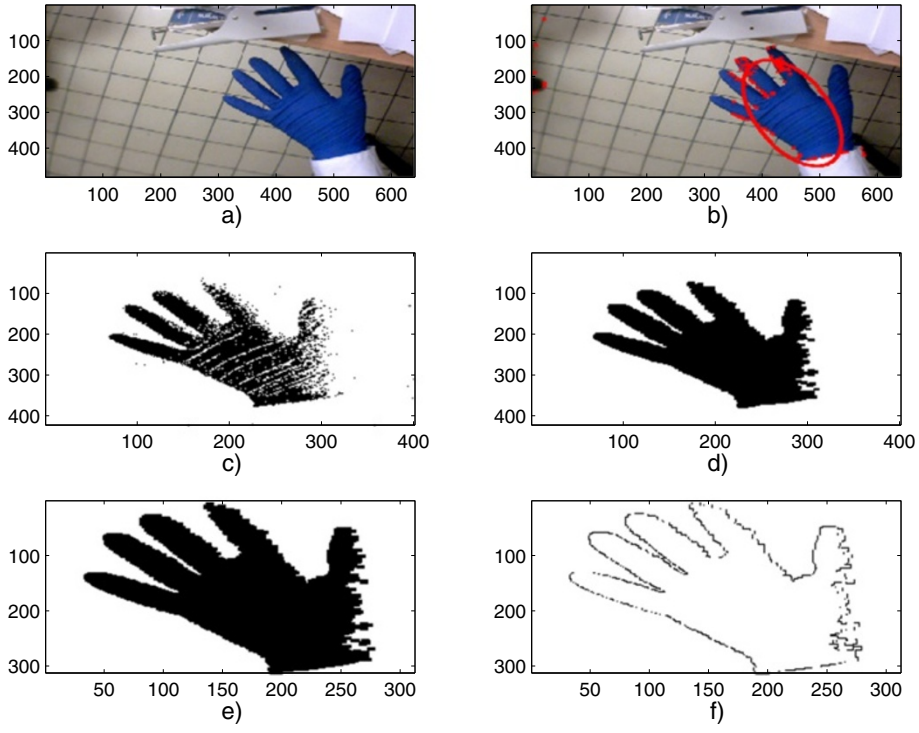

Figure 14 Blue hand 2 and steps of the proposed method. (a) Processed image. **(b)** Moving points, fitting ellipse (indicated by red dots and a red ellipse), and center of mass (indicated by a red square). **(c)** Threshold image I^{Th} in the ROI defined from the fitting ellipse. **(d)** Result obtained after mathematical morphology operations. **(e)** I^{f}-square ROI whose height is the maximum Ferret diameter of the hand. **(f)** I^{c}, obtained from Roberts linear filtering, containing the expected hand contour.

and Q. To generate the signature $\mathbf{Z}$ whose components are $z_{p,i}$, with $p = 1,\ldots,P$, and $i = 1,\ldots,Q$ (see Equation 1) a value of $P = 24$ levels is large enough to get an exclusive signature for each posture and small enough to get a reasonable computational load. A value $Q = 120$ yields a good compromise between computational cost and discriminative capability. While performing the recognition of any test image, the regularization term ϵ in Equation 3 is fixed to $10^{-4}\max(\mathbf{\Lambda})$ where $\max(\cdot)$ stands for maximum value.

Firstly, we illustrate the discriminative capability of the method when one test image is considered. Secondly, we provide a statistical study involving all images of the test database. For each of the 11 classes h, we have used on average $M_h = 250$ images in the learning database and 150 images in the test database. In the test database, for each class, one image out of four contains a hand from either a colored person, or wearing a glove, either black, red, yellow, or blue. The same postures forming 11 classes are used in the comparative studies [3,4,10,40]. The images are taken from the database which is used in [10] and which was created within our laboratory in the past few years. Very similar images are also used in [40] though the hands seem to cover a wider part of the image, thus offering supposedly a better resolution. We enriched the existing database with some additional images of dark-skinned hands and hands wearing black, yellow, blue, and red gloves.

7.2.1 Discriminative capability of the proposed signature

For a given image of the test database, we aim at verifying the discriminative capability of the proposed signature. Let us assume this test image to belong to class $\mathrm{h1} \in \{1,\ldots,11\}$, and let $\mathbf{Z}_{\mathrm{h1}}$ be the signature computed from this image. Let $\bar{\mathbf{Z}}_{\mathrm{h2}}$ be the mean signature obtained from all images of the learning database for class $\mathrm{h2} \in \{1,\ldots,11\}$. The proposed signature will be discriminative if these two conditions hold: $\mathbf{Z}_{\mathrm{h1}}$ should be close to $\bar{\mathbf{Z}}_{\mathrm{h2}}$ when $\mathrm{h1} = \mathrm{h2}$, and far from $\bar{\mathbf{Z}}_{\mathrm{h2}}$ when $\mathrm{h1} \neq \mathrm{h2}$.

We compute the normalized distance $d_{\mathrm{h1,h2}} = \frac{||\mathbf{Z}_{\mathrm{h1}} - \bar{\mathbf{Z}}_{\mathrm{h2}}||}{||\bar{\mathbf{Z}}_{\mathrm{h2}}||}$, where $||\cdot||$ denotes Frobenius norm, for one image of each class h1.

Firstly, let us comment on some features concerning all postures: when $\mathrm{h1} = \mathrm{h2}$, the mean of the distance values is $4.7\ 10^{-1}$; when $\mathrm{h1} \neq \mathrm{h2}$, the mean of the distance values

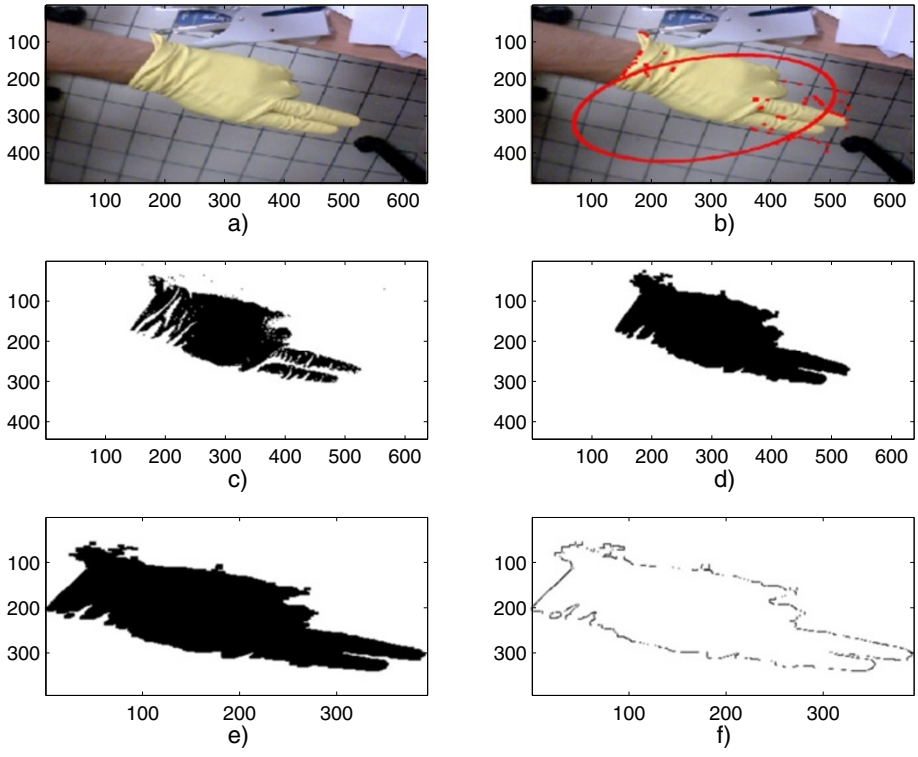

Figure 15 Yellow hand and steps of the proposed method. (a) Processed image. **(b)** Moving points, fitting ellipse (indicated by red dots and a red ellipse), and center of mass (indicated by a red square). **(c)** Threshold image I^{Th} in the ROI defined from the fitting ellipse. **(d)** Result obtained after mathematical morphology operations. **(e)** I^{f}-square ROI whose height is the maximum Ferret diameter of the hand. **(f)** I^{c}, obtained from Roberts linear filtering, containing the expected hand contour.

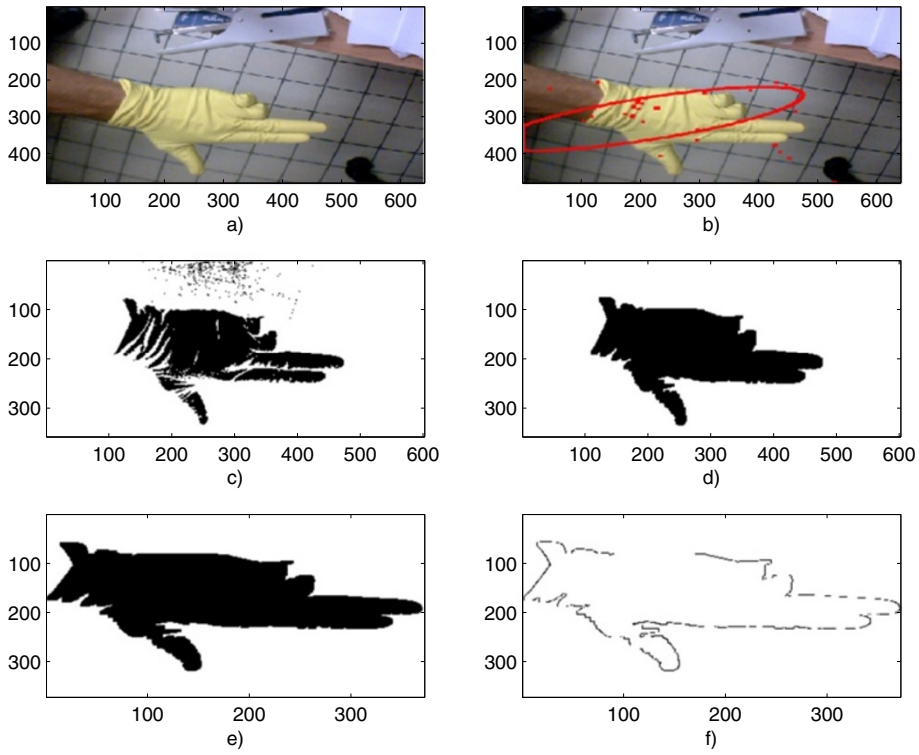

Figure 16 Yellow hand 2 and steps of the proposed method. (a) Processed image. **(b)** Moving points, fitting ellipse (indicated by red dots and a red ellipse), and center of mass (indicated by a red square). **(c)** Threshold image I^{Th} in the ROI defined from the fitting ellipse. **(d)** Result obtained after mathematical morphology operations. **(e)** I^{f}-square ROI whose height is the maximum Ferret diameter of the hand. **(f)** I^{C}, obtained from Roberts linear filtering, containing the expected hand contour.

is 6.5 10^{-1}; and whatever h1 $\in \{1, \ldots, 11\}$, $d_{\mathrm{h1,h2}}$ is minimum when h1 = h2. The distance between two signatures is then the closest when these signatures correspond to images with the same postures.

Secondly, let us focus on specific features: the smallest distance values obtained in the case where h1 $\neq$ h2 are $d_{4,1} = 4.8\ 10^{-1}$ and $d_{8,9} = 4.9\ 10^{-1}$, that is, for couples of visually close postures. The distance $d_{10,6} = 5.8\ 10^{-1}$ is also significantly smaller than the mean value 6.5 10^{-1}. The distance values $d_{4,1}$, $d_{8,9}$, and $d_{10,6} = 5.8\ 10^{-1}$ remain more elevated than the mean value 4.7 10^{-1} obtained when postures are the same.

The reliability of the signature must be proven with a statistical study as will be described in the following.

7.2.2 Statistical study of posture recognition performance

To perform dimensionality reduction (see Subsection 5.1), the number of dominant singular values retained while performing PCA is fixed to $K = 60$. This number, empirically chosen, should be higher than the number of candidate postures (11), and small enough to ensure a reduced computational load. This value of K is also the number of rows and columns in $\mathbf{\Lambda}_h$, a small value which eases its inversion to compute the Mahalanobis distance in Equation 3.

To assess the effectiveness of our proposed technique, we have compared our method with three similar methods: The first method combines Gabor filter, PCA, and support vector machine (SVM) [40]. The second method relies on Fourier descriptors [10,13]. The third one uses the same principles for signature generation [3,4], without however sorting the signature components and differing in the computation of the binary image I which is used as an input for the computation of the contour signature (see [3,4]). This binary image is obtained using YC_bC_r mapping, a threshold applied to the C_r component and mathematical morphology operations. Additionally, in [4], PCA is already used to reduce the dimensionality of the data, but the preselection of candidate postures is not performed with the proposed sphericity, but with a surface criterion.

In Table 1, we depict the results obtained with YC_nbC_r mapping and Fourier coefficients as invariant characteristics. This table shows that Fourier descriptors encounter difficulties with postures 4 (60.8%), 8 (64.8%), and 10 (74.4%). This is due to the unability of Fourier coefficients

Table 1 Confusion matrix (in %, precision 0.1)[13]

	1	2	3	4	5	6	7	8	9	10	11
1	*86.6*	0	0	0	0	0	0	0	0	0	0
2	0	*90.8*	0.4	0.4	0.2	0.2	0.1	0	1.7	0.1	0.1
3	0	0.7	*96.4*	0.5	0.4	1	0.4	0	0.7	0.1	3.3
4	5.5	0	0	*60.8*	0	0.1	0.4	0	0	0	0
5	2.9	1.8	0.5	35.9	*97.8*	0.9	7.8	3.2	4.9	20.2	0.1
6	4.6	0.1	0	0.1	0.3	*94.3*	0.8	0	0.2	2	0
7	0.2	0.4	0.1	0.7	0.5	1.1	*80.6*	8.3	0.3	2.8	0
8	0	0.2	0	0.3	0.3	0.1	1.9	*64.8*	2.8	0.5	0
9	0	5.9	1.7	0.9	0.3	0.4	6	23.2	*88.6*	0.9	0.4
10	0	0.1	0.1	0.3	0	0.2	0.8	0.4	0.2	*73.4*	0
11	0.2	0.2	0.8	0.1	0.1	1.6	1.1	0.1	0.7	0	*96.2*

Obtained with Fourier coefficients and Mahalanobis distance. Italic features are the true-positive recognition rates.

to preserve details: contours are smoothed, and subtle differences such as the presence of one supplementary finger as can be seen to occur between posture 4 and posture 5, and between posture 8 and posture 9, are not detected when Fourier coefficients are used. On the other hand, our method based on signature generation technique offers a 1-pixel resolution and does not encounter such problems.

Table 2 shows the confusion matrix obtained with the signature **Z**, PCA, and Mahalanobis distance [4]. This method is based on YC_bC_r mapping. It clearly shows that this method gives good results, except for posture 4 which is recognized as posture 5 with 11.3 % of the cases, posture 8 is recognized as posture 9 with 25.6 % of the cases, and posture 5 as 4 in 5.5 % of the cases.

The confusion matrix obtained with the method proposed is presented in Table 3.

The comments which result from the confusion matrix of Table 3 obtained with the method proposed in this paper fall into two parts:

Firstly, concerning the true-positive recognition rates, they are more elevated with our method using optical flow, signature **Z**, PCA, and Mahalanobis distance for postures 2, 3, 4, 5, 6, 7, 8, 9, and 10, compared to the case where YC_bC_r mapping is used. The true-positive recognition rates are better in particular for postures 4, 5, and 9, and even much better for posture 8, in which it increases from 64.8% in [10] and 67.4% in [4] to 96.1% with the method proposed in this paper.

Secondly, concerning the false-positive rates, we notice that they are larger than 0 for the couples $\{h1, h2\}$ for which $d_{h1,h2}$ is rather low, that is, for the couples of postures which are visually the closest.

Table 2 Confusion matrix (in %, precision 0.1)

	1	2	3	4	5	6	7	8	9	10	11
1	*97.7*	0	0	0	0	0	0	0	0	0	0
2	0	*100*	0	0	0	0	0	0	0	0	0
3	0	0	*90.8*	0	0	0	2.3	4.7	2.4	0	0
4	0	0	0	*86.4*	5.5	0	0	0	0	0	0
5	2.3	0	2.3	11.3	*91.7*	0	0	0	0	0	0
6	0	0	0	0	0	*95.5*	0	0	0	0	0
7	0	0	0	0	0	0	*93.1*	2.3	0	0	0
8	0	0	0	0	0	0	2.3	*67.4*	2.4	0	0
9	0	0	4.5	2.3	2.8	0	2.3	25.6	*92.8*	2.1	0
10	0	0	0	0	0	4.5	0	0	2.4	*97.9*	0
11	0	0	2.4	0	0	0	0	0	0	0	*100*

Obtained with signature **Z**, PCA, and Mahalanobis distance [4]. Italic features are the true-positive recognition rates.

Table 3 Confusion matrix (in %, precision 0.1)

	1	2	3	4	5	6	7	8	9	10	11
1	*96.3*	0	0	0	0	0	0	0	0	0	0
2	2.4	*100*	0	5.5	1.6	0	0	0	0	0	0
3	0	0	*99.1*	0	0	0	0	0	0.7	0	0
4	1.2	0	0	*92.6*	0	0	0	0	0	0	0
5	0	0	0	1.8	*98.3*	0	0	0	0	0	0
6	0	0	0	0	0	*95.2*	0	0	0	0	0
7	0	0	0	0	0	0	*99.2*	2.3	0	0	0
8	0	0	0	0	0	0	0	*96.1*	1.5	0	0
9	0	0	0.9	0	0	0	0.8	1.5	*97.7*	0	0
10	0	0	0	0	0	4.7	0	0	0	*100*	0
11	0	0	0	0	0	0	0	0	0	0	*100*

Obtained with optical flow, proposed signature, PCA, and Mahalanobis distance. Italic features are the true-positive recognition rates.

We remind also that the results presented in Table 3 were obtained with hands of various colors, and therefore a somehow richer database than in [10] and [4], where only white hands are considered. This attractive performance relies on the quality of the binary images I which are provided to the signature generation method, unlike the YC_bC_r mapping which results in blurring the edges and reduces the contrast between hand surface and background on the C_r channel. The second preprocessing step in Subsection 3.3 consists in thresholding the R,G, and B channels of the RGB color image where the contrast between hand surface and background is increased. The three threshold images are then combined to get a single binary image. Selecting a region of interest with optical flow permits to avoid unexpected noise pixels in this binary image.

These good recognition results are also due to the discriminative capabilities of the signature.

In the following as shown in Table 4 we consider the performance of the proposed method against the other three methods in terms of speed.

Table 4 Proposed and comparative methods, comparison of performances

	Classification method	Speed (frames/s)	System (GHz)	Soft
a	PCA+SVM	4	3.4	C
b	Fourier + Mahalanobis	20	2	C
c	PCA + Mahalanobis	6	3.1	Matlab
d	OF + PCA + Mahalanobis	4	3.1	Matlab + C

a, Gabor-filtered + PCA + SVM [40]; b, Fourier descriptors (FD1) + Mahalanobis; c, YC_bC_r mapping, PCA, and Mahalanobis distance [4]; d, proposed method involving optical flow (OF).

As mentioned in Subsection 7.1, while applying the optical flow, it is necessary to wait 1 10^{-1} s before calculating the velocity vectors for a couple of images t and $t+1$ in order to work in appropriate conditions. The time required by the optical flow component to calculate the velocity vectors and the moving points is much shorter around 3.3 10^{-2} s, which is the time interval between two frames provided by the camera. The following computational times are obtained excluding the optical flow: The off-line methods require 102 s to process the 2,750 images of the learning database, reaching a mean rate of 27 frames/s. For the 1,650 images of the test database, the preprocessing step requires 410 s, that is, 4 images per second; the on-line recognition out of the binary images requires 22 s for the 1,650 images, that is, 27 images per second. Therefore, the preprocessing operations are the most computationally intensive tasks and are the limiting factors. In Table 4 (d), the overall mean rate of 4 frames/s is mentioned. It concerns the on-line recognition step.

The method combining Gabor filter, PCA, and SVM [40] processes 4 frames/s as well (see Table 4 (a)). Fourier descriptors programmed in C++ [13] are faster, namely 20 frames/s (see Table 4 (b)). The method involving YC_bC_r, PCA, and Mahalanobis distance [4] (Table 4 (c)) mapping is faster (6 frames/s) but it exhibits a major drawback as all methods using YC_bC_r mapping: it does not handle colored hands.

Currently, the programmes dedicated to optical flow are not limitative, as they require 1 10^{-1} s per frame, while the subsequent preprocessings and recognition steps require about 0.25 s. The optical flow program is written in C++. This can explain the fact that its computational load is not limitative. We can expect that transferring all our programmes from Matlab® to C++ would decrease the required computational time. We can infer from the good segmentation results obtained

with optical flow that this method may yield in the future a combined exploitation of movement analysis and hand posture recognition for a wide variety of touchless applications.

8 Conclusions

This paper proposes a novel hand posture recognition. The technique has a number of advantages including in invariance to translation, rotation, and scale distortions and can operate with both left and right hands. Before generating the invariant signature of a hand, a binary image containing the hand contour is computed by adapting the concept of optical flow. The proposed method is based on the hand movement and is able to work with colored hands including gloves and employs an elliptic fitting of the moving points to compute a region of interest thereby ensuring the invariance of the signature to scaling and translation. The signature generated from the binary contour image is a sparse matrix, hence, our proposal to apply principal component analysis to reduce the data dimensionality. We also reduce the dimension of the test set through a first rejection test based on a sphericity criterion. Hand posture recognition is eventually performed by computing a Mahalanobis distance between the signature obtained from the test image and preselected reference signatures. The visual results show that despite a complex background, a hand contour is correctly retrieved for various hand colors. Statistical results summarized as a confusion matrix show that the difficult cases of close postures yield a correct recognition result in more than 90% of the cases. Our method offers a good compromise between recognition rate and computational load. In future works, we could determine the exact velocity of the hand with optical flow, to yield a complex set of instructions based on both speed and posture in the frame of a human-machine interaction system.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work was financially supported by the 'Conseil régional Provence Alpes Côte d'Azur' and by the firm 'Intui-Sense Technologies,' to which we are very grateful. The authors would like to thank the anonymous reviewers for their careful reading and their helpful remarks, which have contributed in improving the quality of the paper.

Received: 26 April 2013 Accepted: 22 October 2013
Published: 5 November 2013

References

1. Y Zhu, G Xu, DJ Kriegman, A real-time approach to the spotting, representation, and recognition of hand gestures for human-computer interaction. Comput. Vis. Image Underst. **85**, 189–208 (2002)
2. B Ionescu, D Coquin, P Lambert, V Buzuloiu, Dynamic hand gesture recognition using the skeleton of the hand. EURASIP J Adv. Signal Process. **2005**, 236190 (2005)
3. N Boughnim, J Marot, C Fossati, S Bourennane, Hand posture classification by means of a new contour signature, in *Proceedings of the 14th International Conference, ACIVS 2012,* Brno, Czech Republic, 4–7 September 2012, vol. 7517 2012, pp. 384–394
4. N Boughnim, J Marot, C Fossati, S Bourennane, F Guerault, Fast and improved hand classification using dimensionality reduction and test set reduction, in *Proceedings of ICASSP,* Vancouver, 26–31 May 2013, pp. 1971–1975
5. E Frigerio, M Marcon, S Tubaro, Improving action classification with volumetric data using 3d morphological operators, in *Proceedings of ICASSP,* Vancouver, 26–31 May 2013, pp. 1849–1853
6. M Hu, Visual pattern recognition by moment invariants. IEEE Trans. Inf. Theory. **8**, 179–187 (1962)
7. A Khotanzad, Y Hong, Invariant image recognition by Zernike moments. IEEE-PAMI. **12**(5), 489–497 (1990)
8. E Persoon, K Fu, Shape discrimination using fourier descriptors. IEEE-PAMI. **8**(3), 388–397 (1986)
9. M Charmi, S Derrode, F Ghorbel, Fourier-based geometric shape prior for snakes. Pattern Recognit. Lett. **29**(7), 897–904 (2008)
10. S Bourennane, C Fossati, Comparison of shape descriptors for hand posture recognition in video. Signal, Image Video Process. **6**, 147–157 (2012)
11. J Triesch, C von der Malsburg, Robust classification of hand postures against complex backgrounds, in *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition,* Killington, 14–16 October 1996, pp. 170–175
12. S Marcel, O Bernier, J-E Viallet, D Collobert, Hand gesture recognition using Input/Ouput Hidden Markov Models, in *Proceedings of the 4th International Conference on Automatic Face and Gesture Recognition (AFGR),* Grenoble, March 2000. http://www.idiap.ch/resource/gestures/
13. S Conseil, S Bourennane, L Martin, Comparison of fourier descriptors and Hu moments for hand posture recognition, in *Proceedings of the European Signal Processing Conference (EUSIPCO'07),* Poznan, 3–7 September 2007
14. D Mazumdar, A Talukdar, K Sarma, A colored finger tip-based tracking method for continuous hand gesture recognition. Int. J. Electron. Signals Syst. **3**, 71–75 (2013)
15. M Soriano, B Martinkauppi, S Huovinen, M Laaksonen, Skin detection in video under changing illumination conditions, in *Proceedings of the 15th ICPR,* Barcelona, 3–7 September 2000 vol. 1 2000, pp. 839–842
16. Y Raja, S Kenna M c, S Gong, Tracking colour objects using adaptive mixture models. Image Vision Comput. **17**(3–4), 225–232 (1999)
17. T Yoo, I Oh, A fast algorithm for tracking human faces based on chromatic histograms. Pattern Recognit. Lett. **20**(10), 967–978 (1999)
18. A Ghotka, G Kharate, Hand segmentation techniques to hand gesture recognition for natural human computer interaction. Int. J. Hum. Comput. Interact. **3**, 15–25 (2012)
19. A Chaudhary, J Raheja, K Das, S Raheja, Intelligent approaches to interact with machines using hand gesture recognition in natural way: a survey. Int. J. Comput. Sci. Eng. Surv. **2**, 122–133 (2011)
20. X Zhu, L Yang, W Alex, Segmenting hands of arbitrary color, in *International Conference on Automatic Face and Gesture Recognition,* Grenoble, 28–30 March 2000, pp. 446–453
21. B Horn, B Schunck, Determining optical flow. Artif. Intell. **17**(1–3), 185–203 (1981)
22. B Lucas, T Kanade, An iterative image registration technique with an application to stereo vision, in *Proceedings of the 7th International Joint Conference on Artificial Intelligence,* University of British Columbia, 24–28 August 1981
23. X Papademetris, PN Belhumeur, Estimation of motion boundary location and optical flow using dynamic programming, in *Proceedings of ICIP'96,* Lausanne, 16–19 September 1996, vol. 1 1996, pp. 509–512
24. J Bouguet, Pyramidal implementation of the Lucas Kanade feature tracker: description of the algorithm, technical report. OpenCV documents, Intel Corporation, Microprocessor Research Labs (2000)
25. T Brox, J Malik, Large displacement optical flow: descriptor matching in variational motion estimation. IEEE Trans. PAMI. doi:10.1109/TPAMI.2010.143
26. J Han, F Qi, G Shi, Gradient sparsity for piecewise continuous optical flow estimation, in *Proceedings of ICIP'11,* Brussels, 11–14 September 2011, pp. 2341–2344
27. N Roudel, F Berry, J Serot, L Eck, Hardware implementation of a real time Lucas and Kanade optical flow, in *Proceedings of DASIP'09,* Sophia-Antipolis, France, 22–24 September 2009

28. W Gander, G Golub, R Strebel, Least-squares fitting of circles and ellipses. BIT. **34**, 558–578 (1994)
29. R Xu, O Guida, Comparison of sizing small particles using different technologies. Powder Technol. **132**(2–3), 145–153 (2003)
30. ME Celebi, YA Aslandogan, A comparative study of three moment-based shape descriptors, in *Proceedings of the ITCC'05,* Las Vegas, 4–6 April 2005
31. A Foulonneau, P Charbonnier, F Heitz, Geometric shape priors for region-based active contours, in *International Conference on Imaging Systems and Techniques,* Thessaloniki, 1–2 July 2003, vol. 3 2003, pp. 413–416
32. X Shu, XJ Wu, A novel contour descriptor for 2D shape matching and its application to image retrieval. Image Vision Comput. **29**(2–3), 286–294 (2011)
33. W Gong, J Gonzalez, F Roca, Human action recognition based on estimated weak poses. EURASIP J. Adv. Signal Process. **2012**(162) (2012)
34. J Marot, S Bourennane, Subspace-based and DIRECT algorithms for distorted circular contour estimation. IEEE Trans. Image Process. **16**(9), 2369–2378 (2007)
35. H Jiang, J Marot, C Fossati, S Bourennane, Strongly concave star-shaped contour characterization by algebra tools. Elsevier Signal Process. **92**, 1567–1579 (2012)
36. P Huber, Projection pursuit. Annals of Statistics. **13**(2), 435–475 (1985)
37. A Sharma, K Paliwal, Fast principal component analysis using fixed-point algorithm. Pattern Recognition Letters. **28**, 1151–1155 (2007)
38. A Hyvarinen, E Oja, A fast fixed-point algorithm for independent component analysis. Neural Comput. **9**(7), 1483–1492 (1997)
39. W Kim, S Kim, K Kim, Fast algorithms for binary dilation and erosion using run-length encoding. ETRI. **27**(6), 814–817 (2005)
40. D Huang, WC Hu, SH Chang, Gabor filter-based hand-pose angle estimation for hand gesture recognition under varying illumination. Expert Systems with Applications. **38**, 6031–6042 (2011)

doi:10.1186/1687-6180-2013-167
Cite this article as: Boughnim *et al.*: **Hand posture recognition using jointly optical flow and dimensionality reduction.** *EURASIP Journal on Advances in Signal Processing* 2013 **2013**:167.

Submit your manuscript to a SpringerOpen journal and benefit from:

► Convenient online submission
► Rigorous peer review
► Immediate publication on acceptance
► Open access: articles freely available online
► High visibility within the field
► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com