Hand Posture Recognition with Multiview Descriptors

Jean-François Collumeau[1], Hélène Laurent[1], Bruno Emile[2], and Rémy Leconge[3]

[1] ENSI de Bourges, Laboratoire PRISME 88 bd Lahitolle
18020 Bourges Cedex, France
[2] Laboratoire PRISME, Université d'Orléans, 2 av F. Mitterrand
36000 Châteauroux, France
[3] Laboratoire PRISME, Université d'Orléans, 12 rue de Blois, BP 6744
45067 Orléans cedex 2, France

Abstract. Preservation of asepsis in operating rooms is essential for limiting the contamination of patients by hospital-acquired infections. Strict rules hinder surgeons from interacting directly with any sterile equipement, requiring the intermediary of an assistant or a nurse. Such indirect control may prove itself clumsy and slow up the performed surgery. Gesture-based Human-Computer Interfaces show a promising alternative to assistants and could help surgeons in taking direct control over sterile equipements in the future without jeopardizing asepsis.

This paper presents the experiments we led on hand posture feature selection and the obtained results. State-of-the-art description methods classified in four different categories (i.e. local, semi-local, global and geometric description approaches) have been selected to this end. Their recognition rates when combined with a linear Support Vector Machine classifier are compared while attempting to recognize hand postures issued from an *ad-hoc* database. For each descriptor, we study the effects of removing the background to simulate a segmentation step and the importance of a correct hand framing in the picture. Obtained results show all descriptors benefit to various extents from the segmentation step. Geometric approaches perform best, followed closely by Dalal et al.'s Histogram of Oriented Gradients.

Keywords: Human-Computer Interface, Gesture Recognition, Geometry-Based Hand Description.

1 Introduction

With the development of computer systems and their evergrowing embedded presence into our daily life, the question of convenient and natural types of human-computer-interaction becomes crucial. If user-computer relationships have already evolved in that sense, going from cumbersome text-based command lines to dedicated devices such as mouse or pen, they still remain restrictive. One way to simplify the means of interacting with computers consists in using voice or hand gesture interfaces as people do in their daily life while speaking to one another. Two ways exist to make hand gestures interpreted by computers. The first one relies on the use of extra sensors, such as magnetic ones or data gloves. If these instruments often help in collecting accurate information

J. Blanc-Talon et al. (Eds.): ACIVS 2012, LNCS 7517, pp. 455–466, 2012.

on hand configuration and motion, they also act as a brake upon free movements. The load of cables connected to the computer, induced by this approach, indeed hinders the ease of the user interaction. A less intrusive solution resorts to vision-based systems. Even though it is difficult to intend a generic interface using this technique, this approach has many appealing advantages. The most interesting among these is undoubtly the naturalness of interaction, which results in a much more intuitive communication between human and computers.

Many application domains take interest in gesture interaction, one can quote among others : computer games development, virtual reality, robots control or sign language interpretation [1]. Our work takes place in the specific context of the $CORTECS$ project. This project focuses on intelligent operating rooms (OR) allowing to improve the working conditions of the medical staff including nurses, assistants, surgeons... One objective is to give the operating team the capacity to directly master its environment. Due to asepsis preservation, the interaction with the entire equipment of the OR is restricted for surgeons and assistants who apply for nurse assistance to manipulate non-sterile devices. This results in a paradoxical situation where more and more performing equipments surround the medical staff (as shown in figure 1) while remaining non directly accessible for most of them for fear of contamination. The objective of the project is to make the transition between sterile and non-sterile worlds easier without jeopardizing asepsis preservation.

Fig. 1. Example of operating room design with varied and complex equipments, including mobile lights, fixed or mobile screens and cameras

Due to the OR's noisy environment, voice-controlled systems, as proposed in [2], seem to be less competitive than vision-based ones. First attempts to design remote non-contact OR equipment controls concerned tools for sterile browsing of MRI or CT scan images [3,4], with hand-based commands. The principal drawback of these systems is their low flexibility which forces the surgeon to be positioned in front of the controlled device and consequently to move away from the patient. Within the $CORTECS$ project, the foreseen system should allow the surgeon to interact with various equipments and to choose parameters settings or positions of mobile devices by performing various hand postures. After a preliminary survey conducted with several personnel from various medical branches, it appears that most of time surgeons can

easily free one hand in non-urgent situations during both pre-operative and operative stages. Moreover they are used to wearing medical equipment. Using a camera embedded in the protagonist's kit would therefore not be awkward for the user while allowing the camera to stay at the heart of the operating theater. Thereby we aim at avoiding attention loss from the user by allowing him to remain close to the patient. On the other hand, due to the use of a mobile camera, the system will thus have to be tolerant to disparities in acquisition points of view and in lighting conditions.

We focus in this paper on the hand posture recognition step. Selecting pertinent features is crucial for the whole process' performance. In a preliminary study conducted in [5], we compared the performances of classical object recognition approaches in this specific context where objects, i.e. postures, could only slightly differ in finger positioning. In this article, we complete this study by comparing several global, semi-local, local and geometrical approaches. The corresponding descriptors are presented in section 2. In order to work in more application-dependant conditions, we created an ad-hoc database using surgical materials. Section 3 is dedicated to the presentation of the created database and the test protocol. Relative performances of tested descriptors are compared. The influence of background substraction and object texture are also studied. Indeed these may impact descriptor performance. Finally, first considerations in combining different descriptors are introduced and hand positioning aspects in posture vocabulary definition are considered. Section 4 presents the conclusions and perspectives of this study.

2 Presentation of the Tested Descriptors

In order to characterize an object (here, various hand postures) that can appear at different scales and orientations, an invariant descriptor must be used. The descriptors can be divided into four classes: the global descriptors that work on the entire image, semi-local descriptors that work on a set of sub-images representing cuts of the complete image, local descriptors that combine interest points detection and characterization of the neighborhood of each detected keypoint and geometric descriptors that utilize low level features to express object shape. In the following paragraphs, we detail some descriptors for each class. For this study, these descriptors have been combined with a linear Support Vector Machine *SVM*.

2.1 Global Approach

Zernike Moments (Zer). The Zernike descriptor is among the most used in the literature. It is built from a set of Zernike polynomials. This set is complete and orthonormal in the interior of the unit circle. The Zernike moments have shown their performance in terms of noise resistance and near zero value in redundancy of information. The Zernike moments formulation is given below [6]:

$$A_{mn} = \frac{m+1}{\pi} \sum_{x} \sum_{y} I(x,y)[V_{mn}(x,y)] \tag{1}$$

with $x^2 + y^2 < 1$.

$I(x, y)$ is the pixel gray-level of the image I ; m and n are the values defining the moment order. Zernike polynomials $V_{mn}(x, y)$ are expressed in the radial-polar form:

$$V_{mn}(r, \theta) = \sum_{s=0}^{\frac{m-|n|}{2}} \frac{(-1)^s (m-s)! r^{m-2s}}{s!(\frac{m+|n|}{2} - s)!(\frac{m-|n|}{2} - s)!} e^{-jn\theta} \tag{2}$$

These moments respect translation, scale and rotation invariance. They have been used by [7] for recognizing hand postures in a human-robot interaction context.

Hu Moments (Hu). The seven Hu moments are invariant under translation, rotation and scaling [8] and are calculated from the normalized moments :

$$\mu_{p,q} = \frac{v_{p,q}}{v_{0,0}^{1+(p+q)/2}} \tag{3}$$

with

$$v_{p,q} = \int_{\mathbb{R}^2} x^p y^q I(x + x_0, y + y_0) dx dy \tag{4}$$

(x_0, y_0), centroid of I, is defined by : $x_0 = \frac{m_{1,0}}{m_{0,0}}$ and $y_0 = \frac{m_{0,1}}{m_{0,0}}$. Such moments have been used for hand recognition application [9].

2.2 Semi-local Approach

Histogram of Oriented Gradient Descriptors (HOG). Histogram of Oriented Gradient (HOG) descriptors are features widely used by the object detection and object recognition community. They have been shown to be distinctive and robust under small affine transformations and illumination changes. They are constructed by dividing the image into a dense grid of uniformly spaced cells and then computing the orientation histograms of the image gradient values on each cell. The illumination and contrast changes are taken into account by local normalization of the gradient strengths which requires grouping the cells together into larger, spatially-connected blocks. The HOG descriptor is then the vector of the components of the normalized cell histograms for all the block regions. Dalal et al. [10] have proposed Histogram of Oriented Gradients in the case of human detection. They have also been used for hand posture recognition [11] and gesture recognition [12].

2.3 Local Approach

Scale Invariant Feature Transform ($SIFT$). The Scale Invariant Feature Transform ($SIFT$) is a well known local descriptor created in 1999 by Lowe, allowing to detect and extract features which are invariant to rotation and scale and robust to some variations of illuminations, viewpoints and noise. The $SIFT$ descriptor is computed in four steps [4]. The two first stages correspond to the choice of keypoints, first identifying potential interest points that are scale and rotation invariant and then rejecting the ones

that have low contrast and stability. The two last stages correspond to the descriptor vector computation, assigning one or more orientations to each elected keypoint based on local image gradient directions and using a 4*4 location Cartesian grid to compute the gradient on each location bin on the patch around the keypoint. The $SIFT$ descriptor gives good results in the case of object recognition when it can find relevant keypoints. It has been used by Wang et al. [13] for hand posture recognition with the objective of human-robot interaction.

Speeded Up Robust Feature ($SURF$). Speeded Up Robust Feature $SURF$ was first presented by Bay et al. in 2006. Partly inspired by the $SIFT$ descriptor, $SURF$ also consists in interesting points localization followed by feature descriptors computation. In both cases, the output is a representation of the neighbourhood around an interest point as a descriptor vector. $SURF$ is based on the distribution of first order Haar wavelet responses [14]. One of the principal advantages of $SURF$ is to be several times faster than $SIFT$ while stating to have more discriminative power. It uses the integral images to simplify and to accelerate the computations. Yielding a lower dimensional feature descriptor, it reduces the time for feature computation and matching. In [15], a fast multi-scale feature detection, $SURF$-inspired, and a description method for hand gesture recognition is proposed.

2.4 Geometrical Approach

Varied Form Descriptor (Var). Full reconstruction of the hand is not essential for gesture recognition. Many approaches have instead used the extraction of low-level image measurements for that purpose [16]. Being fairly robust to noise, these characteristics can be extracted quickly. In this approach we created a geometry-based feature vector by gathering simple geometrical characteristics described hereunder :

$$\textbf{Isometric rate} = \frac{hand's\ perimeter^2}{hand's\ area * 4 * \pi} \tag{5}$$

$$\textbf{Lengthening} = \frac{radius\ of\ the\ biggest\ hand\ inscribed\ circle}{radius\ of\ the\ smallest\ hand\ circumscribed\ circle} \tag{6}$$

$$\textbf{Concavity} = \frac{perimeter\ of\ the\ hand's\ convex\ hull}{hand's\ perimeter} \tag{7}$$

$$\textbf{Elongation} = \frac{major\ axis\ of\ the\ hand's\ smallest\ elliptical\ hull}{minor\ axis\ of\ the\ hand's\ smallest\ elliptical\ hull} \tag{8}$$

3 Comparative Study

3.1 Test Database

In order to come closer to operating room conditions, a 6000 pictures database has been acquired by dressing the hand of speakers with surgical gloves. A surgical sheet is used as background and the lighting is provided by a LED dome placed above the operating table. The illuminance measured near the hand varies from 150 to 300 lux.

The prototype used for these acquisitions is presented in figure 2 along with a gray-level image (640 x 480 pixels) extracted from the video captured by the image acquisition system, which is located on the speaker. Because of the important illumination variation of surgical workplaces depending on the performed surgery, colors may fade away or become saturated. Therefore color was regarded as an extra source of uncertainty and a gray-level camera was chosen for the speaker-embedded acquisition system in order to discard it.

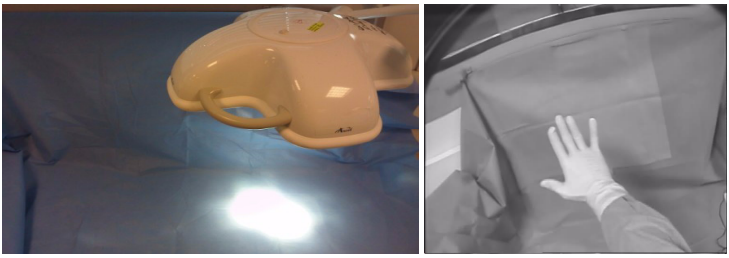

Fig. 2. View of the prototype designed for the database creation using surgical equipments. Picture example acquired through the speaker-embedded camera.

Four speakers have been involved in this experiment. They were asked to reproduce 6 hand postures: 'Y', 'OK', 'Open hand', 'Fist', 'Thumbs up' and 'U'. The postures are presented in figure 3. This posture vocabulary has been selected in order to induce various situations with possible confusions. One can note variations in scale, rotations and lighting conditions. Moreover differences appear between speakers in the vocabulary realization with more or less tensed or spaced fingers. The objectives leading to the creation of such a database are first to test the descriptors' performances to geometrical alterations (i.e. rotations and scaling), and to assert their robustness to both simple and more complicated scenarii (e.g. sparse lighting conditions or inter-speaker posture variability). The final vocabulary will be defined afterwards, drawing lessons from these experiments and being extended or customized to the user's affinity. For each posture, 50 views have been acquired. This was repeated three times for two speakers and twice for the two others. To validate the best hand orientation, the above procedure was realized twice in order to obtain two subsets of 3000 pictures presenting respectively palmar and dorsal aspects of hands. Ground truths were extracted manually for every picture of the database.

3.2 Experiments

The final goal of our work is to enable each surgeon to intervene in the definition of his own vocabulary, resulting in a necessary specific training for each surgeon. As mentioned previously, this results from the various ways different speakers may effect a posture. We decided accordinlgy to train speaker-dependant Support Vector Machine (*SVM*) classifiers. According to Chang and Lin's recommandations on kernel selection [17] in the case of a low ratio between the database's size and the amount of descriptors

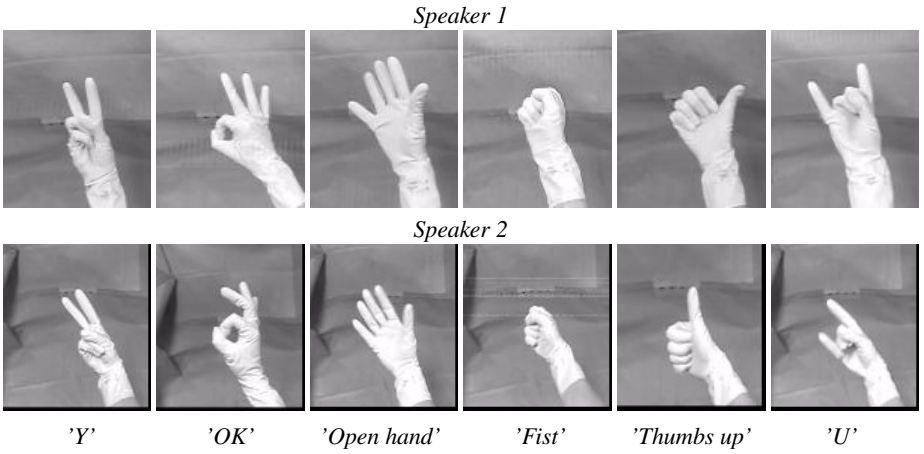

Fig. 3. Examples of the 6 postures constituting the gesture vocabulary for two speakers

involved for describing each of the pictures, we chose a linear kernel. The classifiers were trained using 50 pictures of each posture; 50 other pictures were used to test the classifier performances and compute the recognition rate. Each descriptor was tested using different images as input data: a gray-level image containing both hand and background ($GL\_H\&B$), a gray-level image of the extracted hand on a black background (GL_H) and a binary image containing the mask or contour of the extracted hand (B_M or B_C). Figure 4 presents examples of the images used as input data for the descriptors' computation for a single picture issued from the database.

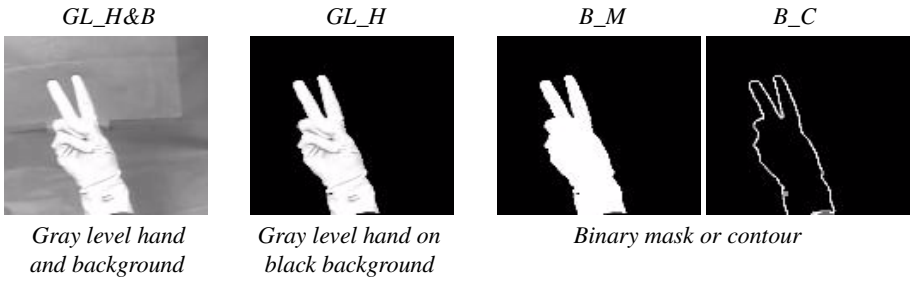

Fig. 4. Input pictures for the computation of the tested descriptors

Table 1 presents the global performances of each descriptor over the whole database corresponding to palmar pictures. The values represent mean recognition rates of the 6 postures for the 4 speakers using different subsets of images for the training and the recognition steps. They are computed averaging sixteen different tests.

Best results are obtained using extracted masks or contours of hands as input images for the descriptors' computation, emphasizing the need for a segmentation step to remove the background in order to achieve satisfactory results. One can observe that the geometrical approach provides an interesting compromise between complexity and

Table 1. Mean recognition rates obtained over the 4 speakers with images presenting palmar aspect

Palmar aspect			
Recognition rates (%)	Gray-level hand and background	Gray-level hand on black background	Binary object
Zer	21,1	24,9	25,6
Hu	19,7	52,5	**68,1**
HOG	33,2	**44,3**	38,2
SIFT	58,1	60,3	63,5
SURF	51,5	60,1	**66,8**
Var	-	-	**76,4**

performance. Even though it involves the preliminary extraction of the object's mask or contour, this approach supplies the best mean recognition rate while requiring very few features. Also depicting the object's geometry, Hu moments arrive second and outperform *HOG* and *SIFT/SURF* features. Both features from the semi-local and local approaches probably suffer in these tests from the relatively large background of the tested images. This can result in the consideration of a few erroneous interest points by *SIFT/SURF* keypoint detectors when full images are taken as input, hence characterizing partly the hand and partly the background. Such confused characterization may lead to confusions between the different hand postures and hinder recognition. Moreover, the poor performances obtained by the local approach may also result from the relative similarity of the objects (i.e. the hand in different configurations) considered in our application. Indeed, many hand postures differ only slightly and share several to most of their features.

In order to understand better the influence of the background presence, we realized the same tests using images restricted to the nearest hand environment. Figure 5 presents examples of the images used as input data for the descriptors' computation when considering reduced background. One can notice in table 2 that the semi-local approach is effectively influenced by this parameter and will therefore by highly dependant on the hand detection step. This is also the case for the Zernike moments and *HOG* descriptor. On the contrary, Hu moments and local features obtain similar results with and with-

Fig. 5. Input images for the computation of the tested descriptors - Reduced background

Table 2. Mean recognition rates obtained over the 4 speakers with images presenting palmar aspects - Reduced background

Palmar aspect			
Recognition rates (%)	Gray-level hand and background	Gray-level hand on black background	Binary object
Zer	41,5	35,5	43,3
Hu	27,6	52,1	**68,2**
HOG	56,9	68,5	**70,4**
SIFT	63,4	64,7	58,5
SURF	57,0	57,4	**67,9**
Var	-	-	**76,4**

out reducing the background. As expected, the geometrical approach's results remain stable.

We were also interested in studying the influence of the point of view under which the hand is seen. In order to check whether using a frontal or a dorsal view of the hand influences the descriptors performances, we reproduced the comparative study conducted with reduced background, on a second set of images presenting dorsal aspects of hand. The corresponding results are gathered in table 3. In comparison to the ones presented in table 2, one can observe that results corresponding to dorsal views are generally similar or poorer than the ones on palmar views. This comment is particularly true when considering gray-level images as input data for the descriptors' computation. This can be explained by the presence of particular folds on glove appearance for some of the selected hand postures. Each palmar posture including tucked fingers will present typical areas with highly structured gray-level contrasts likely to represent potential significant zones or points of interest, whereas dorsal pictures lack these areas. Best results being obtained on palmar images, further developments have been realized considering this situation which also corresponds to more ergonomic-friendly positions.

Table 3. Mean recognition rates obtained over the 4 speakers with images presenting dorsal aspects - Reduced background

Dorsal aspect			
Recognition rates (%)	Gray-level hand and background	Gray-level hand on black background	Binary object
Zer	38,5	34,8	42,9
Hu	22.8	47.3	**69,8**
HoG	50.3	62.2	**68.7**
SIFT	55,0	**61,0**	48,0
SURF	53,8	56,1	**61,0**
Var	-	-	**76,0**

Table 4. Example of confusion matrix obtained for speaker 1, with Hu descriptor, on images with reduced background

Palmar aspect						
Confusion matrix *Speaker 1* Hu *descriptor*	'Y'	'OK'	'Open hand'	'Fist'	'Thumbs up'	'U'
'Y'	**50**	1	0	9	0	12
'OK'	0	**43**	0	0	0	0
'Open hand'	0	2	**50**	0	0	0
'Fist'	0	0	0	**41**	0	0
'Thumbs up'	0	0	0	0	**1**	0
'U'	0	4	0	0	49	**38**

Table 5. Example of confusion matrix obtained for speaker 1, with Var descriptor, on images with reduced background

Palmar aspect						
Confusion matrix *Speaker 1* Var *descriptor*	'Y'	'OK'	'Open hand'	'Fist'	'Thumbs up'	'U'
'Y'	**6**	0	0	0	0	6
'OK'	0	**44**	13	0	0	1
'Open hand'	0	0	**37**	0	0	0
'Fist'	2	0	0	**50**	1	0
'Thumbs up'	0	0	0	0	**49**	1
'U'	42	6	0	0	0	**42**

Table 6. Example of confusion matrix obtained for speaker 1, with Hu-Var combination, on images with reduced background

Palmar aspect						
Confusion matrix *Speaker 1* $Hu - Var$ *combination*	'Y'	'OK'	'Open hand'	'Fist'	'Thumbs up'	'U'
'Y'	**48**	1	0	2	0	10
'OK'	0	**42**	1	0	0	0
'Open hand'	0	4	**49**	0	0	0
'Fist'	2	0	0	**48**	0	0
'Thumbs up'	0	0	0	0	**44**	2
'U'	0	3	0	0	6	**38**

Finally, a promising lead would consist in joining various descriptors together in order to improve the recognition rates through descriptor cooperation. Tables 4 and 5 present the confusion matrices obtained for speaker 1 with respectively Hu and Var

descriptors considered individually. We still consider here images with reduced background. 50 images were used as a training set while a second set of 50 images was used for the recognition characterization. Columns correspond to the real hand posture, while rows correspond to the SVM classification output. Confusions occur between postures using both *Hu* and *Var* approaches, but the two descriptors misclassify different postures. In order to get some idea on the benefit which could be expected from descriptors combinations, we present in table 6 the confusion matrix obtained on the same data combining Hu and Var descriptors. An interesting gap in the global performances can be noticed – 89.7% recognition rate when using the combination versus 74.3% and 76.0% using respectively *Hu* and *Var* independently – thanks to the removal of many confusions. Remaining misclassifications are mainly due to the relative similarity of 'Y' and 'U' postures when seen under specific angles. To better exploit the capacity of existing features, further investigations on descriptors' complementarity in differentiating postures will be conducted. If performed for every speaker, such studies may in addition help adapting the process and the chosen vocabulary to each users' specificity.

4 Conclusion

In this paper, we introduced the non-contact Human-Computer Interface part of the $CORTECS$ project. An extensive *ad-hoc*-created picture database including various speakers effecting multiple hand postures is used as a benchmark for computing geometric, global, semi-global and local descriptors' performances when associated to a linear *SVM* and comparing them. Geometrical approach *Var* and geometry-based global approach Hu moments perform best with respectively 76.4% and 68.2% recognition rates but require a segmentation step prior to their computation. They are followed by keypoint-based local methods (*SIFT, SURF*) whose performance is little enhanced by the segmentation step. *HOG* proved to be especially dependant on the correct framing of the hand, performing poorly when facing a large background-enclosed hand but achieving second best recognition rate (70.4%) when the hand is well-framed. Although less improved than Hu moments by the segmentation step, *HOG*'s performance nevertheless suffers from its lack. Zernike moments come last with a less than 25% recognition rate.

These results outline the worthiness of simple, geometrical descriptors for describing a single object, namely the user's hand, displayed in various configurations. Predominance of such descriptors conveying the hands' shape will therefore focus future research on descriptors whose relevance have been established when dealing with shapes. Descriptor cooperation showed promising prospects and will be studied in greater depth in future works. To this end, data fusion between such descriptors through various means, like *a priori* descriptor concatenation or confidence-based *a posteriori* label decision fusion, will be investigated.

Acknowledgment. The authors would like to thank the Regional Council of the Centre and the French Industry Ministry for their financial support within the framework of the Cortecs project through the Competitiveness Pole S2E2.

References

1. Murthy, G.R.S., Jadon, R.S.: A review of vision based hand gestures recognition. International Journal of Information Technology and Knowledge Management 2(2), 405–410 (2009)
2. Hansen, T.R., Bardram, J.E.: ActiveTheatre - a Collaborative, Event-based Capture and Access System for the Operating Theatre. In: Beigl, M., Intille, S.S., Rekimoto, J., Tokuda, H. (eds.) UbiComp 2005. LNCS, vol. 3660, pp. 375–392. Springer, Heidelberg (2005)
3. Wachs, J.P., Stern, H.I., Edan, Y., Gillam, M., Handler, J., Feied, C., Smith, M.: A gesture-based tool for sterile browsing of radiology images. Journal of the American Medical Informatics Association 15(3), 321–323 (2008)
4. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
5. Collumeau, J.-F., Leconge, R., Emile, B., Laurent, H.: Hand-gesture recognition: comparative study of global, semi-local and local approaches. In: Proc. International Symposium on Image and Signal Processing and Analysis (ISPA), pp. 247–252 (2011)
6. Khotanzad, A., Hong, H.: Invariant image recognition by Zernike moments. IEEE Transactions on Pattern Analysis and Machine Intelligent 12(5), 489–497 (1990)
7. Gu, L., Su, J.: Natural hand posture recognition based on Zernike moments and hierarchical classifier. In: Proc. IEEE International Conference on Robotics and Automation (ICRA), pp. 3088–3093 (2008)
8. Hu, M.K.: Visual pattern recognition by moments invariants. IEEE Transaction on Information Theory 8, 179–187 (1962)
9. Gowtham, P.N.V.S.: An Interactive Hand Gesture Recognition System on the Beagle Board. In: International Proceedings of Computer Science and Information Technology, vol. 20, pp. 113–118 (2011)
10. Dalal, N., Triggs, B., Schmid, C.: Human Detection Using Oriented Histograms of Flow and Appearance. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 428–441. Springer, Heidelberg (2006)
11. Fang, Y., Cheng, J., Wang, J., Wang, K., Liu, J., Lu, H.: Hand posture recognition with co-training. In: Proc. International Conference on Pattern Recognition (ICPR) (2008)
12. Kaâniche, M.B., Brémond, F.: Tracking HOG descriptors for gesture recognition. In: Proc. IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) (2009)
13. Wang, C.-C., Wang, K.-C.: Hand posture recognition using adaboost with SIFT for human robot interaction. In: Proc. International Conference on Advanced Robotics (ICAR) (2007)
14. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: SURF: Speeded-Up Robust Features. In: Computer Vision and Image Understanding (CVIU), vol. 110(3), pp. 346–359 (2008)
15. Fang, Y., Cheng, J., Wang, K., Lu, H.: Hand gesture recognition using fast multi-scale analysis. In: Proc. International Conference on Image and Graphics (ICIG), pp. 694–698 (2007)
16. New, J.R., Hasanbelliu, E., Aguilar, M.: Facilitating User Interaction with Complex Systems via Hand Gesture Recognition. In: Proc. Southeastern ACM Conference (2003)
17. Chang, C.-C., Lin, C.-J.: LIBSVM: a library for support vector machines (2001) Software, `http://www.csie.ntu.edu.tw/cjlin/libsvm`