BULETINUL INSTITUTULUI POLITEHNIC DIN IAŞI
Publicat de
Universitatea Tehnică „Gheorghe Asachi” din Iaşi
Tomul LVII (LXI), Fasc. 6, 2011
Secţia
ELECTROTEHNICĂ. ENERGETICĂ. ELECTRONICĂ

FEATURE DETECTION FOR GESTURE ANALYSIS IN A SPARSE FRAMEWORK

BY

GEORGIANA SIMION*

“Politehnica” University of Timişoara

Received, May 31, 2011
Accepted for publication: July 22, 2011

Abstract. Any object recognition approach has a feature extraction/ selection stage. Features should be carefully selected because there are used to object representation. The main purpose of this work is to find good sparse features which can be used further to recognize hand gestures. The proposed features are edges, lines and patches extracted with Kadir and Brady detector. These simple features can be used to define a large set of hand postures.

Key words: edge; sparse; hand gesture.

1. Introduction

The task of selecting relevant features in classification problems can be viewed as one of the most fundamental problems in the field of machine learning. The ideal features are not affected by occlusion and clutter, there are invariant (or covariant), there are also robust, which means that noise, blur, discretization, compression, etc., do not have a big impact on them. From the distinctive point of view individual features can be matched to a large database of objects; from the quantitative point of view many features can be generated for even small objcts and offer a precise localization.

* *e-mail*: Georgiana.simion@etc.upt.ro

There is a wide variety of interest points and corners detectors in literature. Schmid & Bauckhage (2000) have proved that they can be divided into three categories: contour based, intensity based and parametric model based methods.

Contour based methods have existed for a long time (Asada, 1986; Mokhtarian, 1998) and first extract contours, then search for maximal curvature or inflexion points along the contour chains, or do some polygonal approximation and then search for intersection points. *Intensity based* methods (Tomasi, 1991; Smith, 1997; Harris, 1988) compute a measure that indicates the presence of an interest point directly from the grey values. *Parametric model* methods fit a parametric intensity model to the signal. Usually these methods provide sub-pixel accuracy, but unfortunately are limited to specific types of interest points.

The paper is organized as follows: the second section introduces related works in feature selection with regards to the hand gesture; the third section describes the proposed features to be use in hand gesture recognition; the fourth section includes results regarding two different data base and the conclusions.

2. Related Work

There are a multitude of features used for hand gesture recognition. Hand silhouette is among the simplest, yet most frequently used features. Silhouettes are easily extracted from local hand and arm images in the restrictive background setups. By silhouette it is meaning the outline of the hand provided by segmentation algorithms, or equivalently the partitioning of the input image into object and background pixels. In the case of complex backgrounds, techniques that employ colour histogram analyses can be used. In order to estimate the hand pose the position and orientation of the hand, fingertip locations, and finger orientation from the images are extracted. The center of gravity (or centroid) of the silhouette (Sato & Koik, 2001) is one choice, but it may not be very stable relative to the silhouette shape due to its dependence on the finger positions. The point having the maximum distance to the closest boundary edge (Mo & Neumann, 2005; Oka & Koike, 2002; Abe & Ozawa, 2000), has been argued to be more stable under changes in silhouette shape.

A frequently used feature in gesture analysis is the fingertip. Fingertip detection can be handled by correlation techniques using a circular mask (Oka & Koike, 2002; Koike & Kobayashi, 2001; Letessier, 2004), which provides rotation invariance, or fingertip templates extracted from real images (O'Hagan 1997). Another common method to detect the fingertips and the palm-finger intersections (Segen, 1998; O'Hagan & Rougeaux, 2002; Malik, 2004) is to use the curvature local maxima on the boundary of the silhouette. For the curvature-based methods, in case of noisy silhouette contours, the sensitivity to noise can be an issue. Mo & Neumann (2005) have utilized, a more reliable algorithm based on the distance of the contour points to the hand position. The local maximum of the distance between the hand position and farthest boundary point at each direction gives the fingertip locations. The direction of the principal axis of the silhouettes (Pavlovic & Huang, 1996; Utsumi, 1999) can be used to estimate the finger or 2-D hand orientation. All these features can be tracked across frames to increase computation speed and robustness using Kalman filters (Mo & Neumann, 2005; Martin & Crowley, 1998) or heuristics that determine search windows in the image (O'Hagan & Rougeaux, 2002; Maggioni, 1998) based on previous feature locations or rough planar hand models. The low computational complexity of these methods enables real-time implementations using conventional hardware but their accuracy and robustness are arguable. These methods rely on high quality segmentation lowering their chance of being applied on highly cluttered backgrounds. Failures can be expected in some cases such as two fingers touching each other or out of plane hand rotations.

Jennings *et al*. (1999) have used a more elaborate method by tracking the features directly in 3-D using 3-D models; their method has employed range images, colour, and edge features extracted from multiple cameras to track the index finger in a pointing gesture. Very robust tracking results over cluttered and moving.

3. Finding Good Features for a Sparse Framework

It is very important which features are used for the sparse hand posture representation. The potential benefits of feature selection include, first and foremost, better accuracy of the inference engine and improved scalability. Secondary benefits include better data visualization and understanding, reduced measurement and storage requirements, and reduced training and inference time.

The first question is how hand can be represented in order to be decided which image locations had to be captured and which to dispose of. The main idea is that each hand posture can be described by: the V shapes between the fingers when these are apart, the curve shapes which correspond to the fingertips and the straight lines for the finger length. Each hand pose can be defined as a combination of these shapes. Based on the number of V shapes, curves and lines and based on the relations among them, the hand pose can be recognized. The second question is how these relevant image regions can be represented.

In this paper the Canny edge detector is proposed to extract the hand contour. First the RGB image is converted in to a gray scale image. In order to detect all useful parts of the hand contour it is necessary to use a high sensibility threshold. In this last case the details from the hand are also detected. These details are not useful for the next processing steps. In order to discard them, a function which counts the number of pixels from a connected object is used. The result can be seen in Fig. 1 *d*.

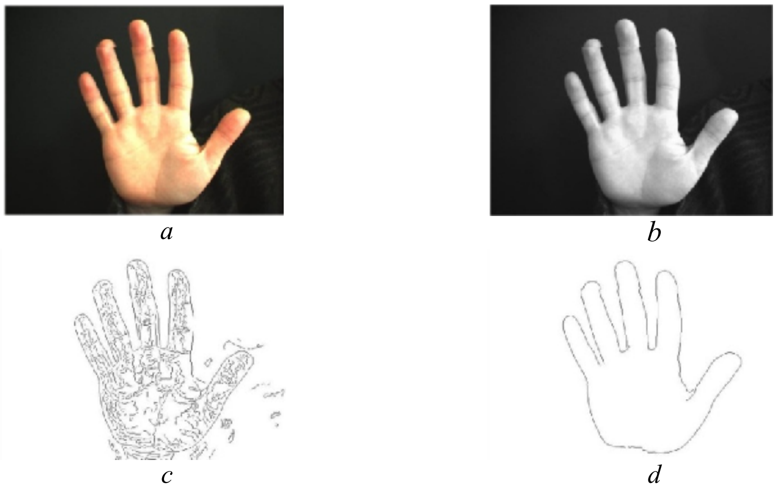

Fig. 1 – Good hand contour: *a* – RGB image; *b* – gray scale image; *c* – Canny 0.05 threshold; *d* – the hand contour after being processed.

Next step was to detect the straight lines for the finger length. Based on the hand contour obtained in the previous stage and using the Hough transform (Gonzalez, 1993) we obtain the results which are presented in Fig. 2. The Hough transform is designed to detect lines, using the parametric representation of a line

$$\rho = x \cos \theta + y \sin \theta. \tag{1}$$

The variable ρ is the distance from the origin to the line along a vector perpendicular to the line; θ is the angle between the x-axis and this vector. The hough function which can be found in Matlab implements the Standard Hough

Fig. 2 –Straight lines detected on the hand contour.

Transform (SHT). The hough function generates a parameter space matrix whose rows and columns correspond to these ρ and θ values, respectively. The houghpeaks function finds peak values in this space, which represent potential lines in the input image. The houghlines function finds the endpoints of the line segments corresponding to peaks in the Hough transform, and it automatically fills in small gaps.

Kadir and Brady detector (2001), suitable for finding circular structures, is proposed to extract the fingertips and the V shapes between the fingers when these are apart. The Kadir and Brady detector searches for scale localized features with high entropy, with the constraint that the scale is isotropic. The algorithm is suitable for finding circular structures. The algorithm generates a space sparsely populated with scalar saliency values. For each pixel location and for each scale value between a minimum and a maximum the local descriptors value is measured within a window; them the PDF from this is estimated and the local entropy is calculated. The scale which conducts to the peaked entropy is selected. A final clustering of the candidates in scale space does then yield a set of interest points. This method finds regions that are salient over both location and scale.

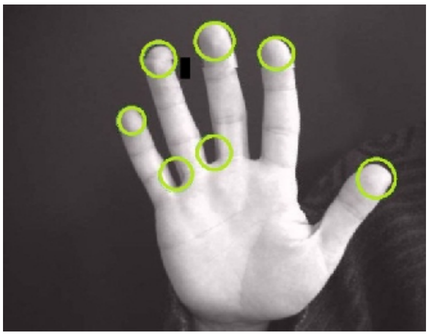

Fig. 3 – Results for the Kadir and Brady detector.

4. Results and Conclusions

The proposed features were extracted for different hand postures from Massay database. These pictures consist in hand postures indoor taken, while artificial light was added. In these images one, two, three, four or five fingers are shown. The size of these pictures is 640 × 480. The results are presented in the Figs. 4,…,6. The threshold for Canny edge detector was set to 0.1. The number of pixels from an area is greater than 150 pixels. The parameters for the Hough transform are: 'RhoResolution' which was set on 5 and 'Theta' set on -90:5:89.5. The parameter of houghpeaks function – numpeaks is a scalar value that specifies the maximum number of peaks to identify and in these experiments it was set on 20. For houghlines function the parameters are: 'FillGap', which is a positive real scalar value that specifies the distance between two line segments associated with the same Hough transform bin – was set on 15; and 'MinLength', a positive real scalar value that specifies whether merged lines should be kept or discarded – was set on 60. Lines shorter than the value specified are discarded. Regarding Kadir and Brady detector parameters, the start scale was set on 15, the stop scale was set on 20 and the threshold on saliency values was set on 0.9.

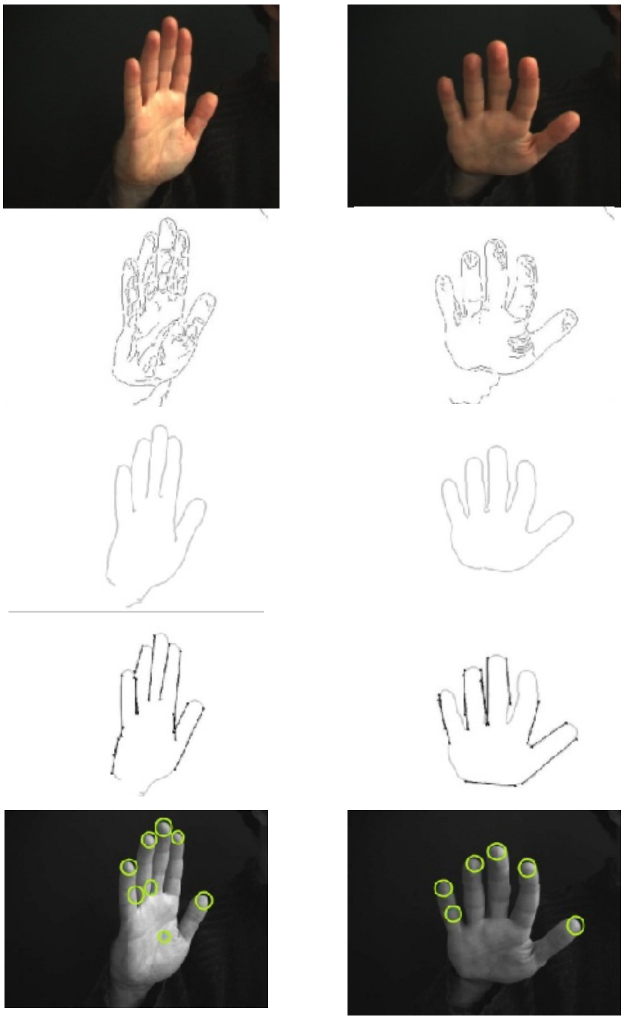

Fig. 4 – Results for five fingers hand posture: the first row represents the original RGB image, the second row represents the result for Canny with threshold set to 0.1, the third row represents the result for the hand contour after being processed, the fourth row represents the lines detected on the hand contour with Hough transform, and the last row represents the result for Kadir and Brady detector.

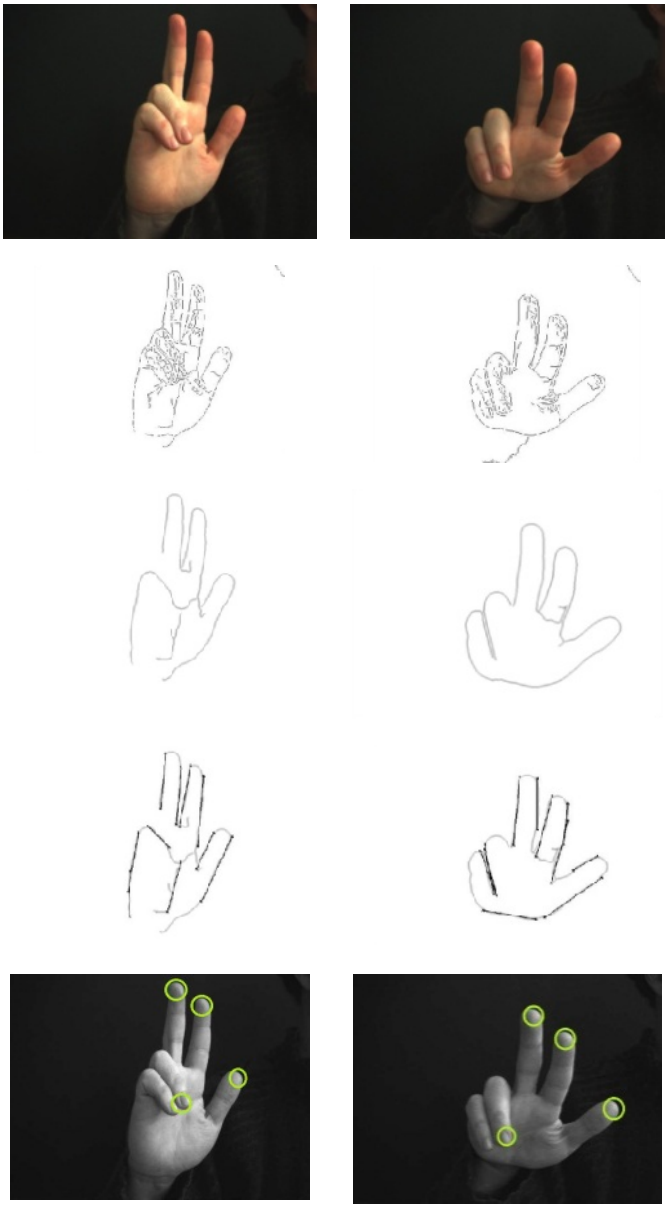

Fig. 5 –Results for three fingers hand posture: the first row represents the original RGB image, the second row represents the result for Canny with threshold set to 0.1, the third row represents the result for the hand contour after being processed, the fourth row represents the lines detected on the hand contour with Hough transform, and the last row represents the result for Kadir and Brady detector.

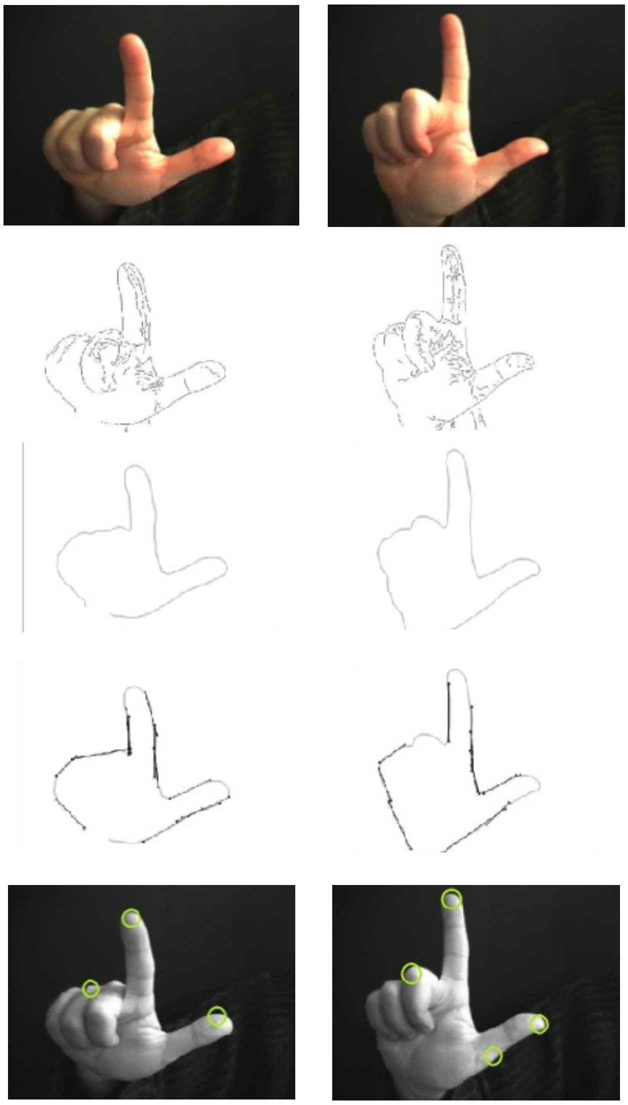

Fig. 6 – Results for two fingers hand posture: the first row represents the original RGB image, the second row represents the result for Canny with threshold set to 0.1, the third row represents the result for the hand contour after being processed, the fourth row represents the lines detected on the hand contour with Hough transform, and the last row represents the result for Kadir and Brady detector.

In this work several sparse features are extracted and proved to be good features for hand gesture recognition in a sparse framework. Using edges, lines and fingertips extracted with Kadir and Brady detector one can define a large set of hand postures. The hand posture is determined by its constituent parts – fingers/fingertips and the relationships between them. Based on the principle of compositionality each hand posture can be recognized from simple parts, like the Lego parts, which are not so varied but can be combining in a flexible way generating objects from houses to cars and planes.

Acknowledgments. This paper was supported by the project "Development and Support of Multidisciplinary Postdoctoral Programmes in Major Technical Areas of National Strategy of Research – Development – Innovation" 4D-POSTDOC, contract no. POSDRU/89/1.5/S/52603, project co-funded by the European Social Fund through Sectoral Operational Programme Human Resources Development 2007-2013.

REFERENCES

Abe K., Saito H., Ozawa S. , *3-D Drawing System Via Hand Motion Recognition from Two Cameras.* IEEE Internat. Conf. on Syst., Man, a. Cybernet., Oct. 8-11, Nashville, TN, USA, 2000, **2**, 840-845.

Asada H., Brady M. , *The Curvature Primal Sketch.* IEEE Trans. on Pattern Anal. a. Machine Intell., **8**, *1*, 2-14 (1986).

Gonzalez R.C., Woods E.R., *Digital Image Processing.* Prentice Hall, NY, 1993.

Harris C.S.M., *A Combined Corner and Edge Detector*. Alvey Vision Conf., 1988, 147-151.

Jennings C., *Robust Finger Tracking with Multiple Cameras.* Internat. Workshop on Recogn., Anal. a. Tracking of Faces a. Gest. in Real-Time Syst., Sept. 26-27, Corfu, Greece, 199, 152-160.

Kadir T., Brady M., *Saliency, Scale and Image Description.* Internat. J. of Comp. Vision, **42**, *2*, 83-105 (2001).

Koike H., Sato Y., Kobayashi Y., *Integrating Paper and Digital Information on Enhanced Desk: a Method for Realtime Finger Tracking on an Augmented Desk System.* ACM Trans. on Comp.-Hum. Inter., **8**, *4*, 307-322 (2001).

Letessier J. , Berard F., *Visual Tracking of Bare Fingers for Interactive Surfaces.* 17th Ann. ACM Symp. on User Interface Software a. Technol., Oct. 24-27, Santa Fe, New Mexico, USA, 2004, 119-122.

Maggioni C., Kammerer B., *Gesture Computer – History, Design and Applications.* In R. Cipolla, A. Pentland (Eds.), *Computer Vision for Human–Machine Interaction*, Cambridge, 1998, 312-325.

Malik S., Laszlo J., *Visual Touchpad: a Two-Handed Gestural Input Device.* 6th Internat. Conf. on Multimodal Interfaces, October 13-15, State College, USA, 2004, 289-296.

Martin J., Devin V., Crowley J.L. , *Active Hand Tracking.* 3rd. Internat. Conf. on Face & Gest. Recogn., IEEE Comp. Soc., April 14-16, Nara, Japan, 1998.

Mo Z. , Lewis J.P., Neumann U. , *Smartcanvas: a Gesture-Driven Intelligent Drawing Desk System.* The 10th Internat. Conf. on Intell. User Interfaces, ACM Press, New York, 2005, 239-243.

Mokhtarian F., Suomela R., *Robust Image Corner Detection through Curvature Scale Space.* IEEE Trans. on Pattern Anal. a. Machine Intell., **20**, *12*, 1376-1381 (1998).

O'Hagan R., Zelinsky A., *Finger Track – A Robust and Real-Time Gesture Interface.* 10th Austral. Joint Conf. on Artif. Intell., Nov. 30, Perth, Australia, 1997, 475-484.

O'Hagan R.G. , Zelinsky A., Rougeaux S, *Visual Gesture Interfaces for Virtual Environments.* Interacting with Comp., **14**, 231-250 (2002).
Oka K., Sato Y., Koike H., *Real-Time Tracking of Multiple Fingertips and Gesture Recognition for Augmented Desk Interface Systems.* Fifth IEEE Internat. Conf. on Autom. Face a. Gest. Recogn., IEEE Comp. Soc., Washington, USA, May 21, 2002, 429-434.
Pavlovic V.I., Sharma R., Huang T.S., *Invited Speech: Gestural Interface to a Visual Computing Environment for Molecular Biologists.* 2nd Internat. Conf. on Autom. Face a. Gesture Recogn., IEEE Comp. Soc., Oct. 13-16, Killington, Vermont, 1996.
Sato Y. , Saito M., Koik H., *Real-Time Input of 3-D Pose and Gestures of a User's Hand and its Applications for HCI.* IEEE Virtual Reality Conf., IEEE Comp. Soc., March 13-17, Yokohama, Japan, 2001.
Schmid C., Mohr R., Bauckhage C., *Evaluation of Interest Point Detectors.* Internat. J. of Comp. Vision, **37**, *2*, 151-172 (2000).
Segen J., Kumar S., *Gesture VR: Vision-Based 3-D Hand Interface for Spatial Interaction*. Sixth ACM Internat. Conf. on Multimedia, Sept. 12-16, Bristol, England, 1998, 455-464.
Smith S.M., Brady J.M., *SUSAN–A New Approach to Low Level Image Processing.* Internat. J. of Comp. Vision, **23**, *1*, 45-78 (1997).
Tomasi C., Kanade T., *Detection and Tracking of Point Features.* Techn. Rep., Carnegie Mellon Univ., **CMU-CS-91-132**, 1991.
Utsumi A., Ohya J., *Multiple-Hand-Gesture Tracking Using Multiple Cameras.* IEEE Comp. Soc. Conf. on Comp. Vision a. Pattern Recogn., June 22-23, Ft Collins Colorado, USA, 1999, 473-478.

DETECȚIA CARACTERISTICILOR PENTRU ANALIZA GESTURILOR ÎNTR-UN CADRU SPARSE

(Rezumat)

Sunt extrase cateva caracteristici din imagini ce conțin gesturi făcute cu mâna, dovedindu-se că sunt caracteristici bune în a fi utilizate pentru recunoaşterea de gesturi. Folosind contururi, linii de pe contur şi vârfuri de degete extrase cu detectorul propus de Kadir şi Brady, se pot defini o serie de gesturi. Fiecare gest poate fi definit în funcție de numărul de linii şi vârfuri de degete pe baza relațiilor dintre ele, asemenea pieselor dintr-un Lego – nu foarte variate ca structură, dar având posibilitatea de a le combina într-o manieră flexibilă, se pot genera diferite obiecte de la clădiri la avioane.