Buletinul Ştiinţific al Universităţii "Politehnica" din Timişoara

Seria ELECTRONICĂ şi TELECOMUNICAŢII
TRANSACTIONS on ELECTRONICS and COMMUNICATIONS

Tom 53(67), Fascicola 2, 2008

Hand Edge Detection for Gesture Analysis in a Sparse Framework.

Georgiana Simion[1], Vasile Gui[2], Marius Otesteanu[3], Daniel Popa[4], Ciprian David[5]

Abstract – In this work a specialized technique to detect hand edges is presented. An optimized HSV to BW space transformation have been searched, with the objective of maximizing correct hand contour detection. The hands edges have been better extracted with our specific transformation and false hands edges have been removed.

Keywords: hand gestures, edge, sparse

I. INTRODUCTION

Humans perform various gestures in everyday life. Hand's gestures are one of the most familiar gestures. Gesture recognition is an area of active current research in computer vision. There are different approaches for hand gesture analysis. Gesture can be classified as static and dynamic gestures. Our purpose is to recognize static hand gestures using a compositional technique.

Sparse representations are compositional techniques.

Sparse techniques have very good results for object categorization when we deal with complex applications [1]. Patches, atoms, salient points, interest points , edges [2, 3 ,4, 5 ,6] are representatively features for sparse representation.

It has been proven that edges have a very important role for humans' sight. Humans are able to recognize people from sketches because this contains useful information for our brain view-processor. In this work has been optimized edges detection for the sparse analysis. Our goal is to detect better the hand in order to recognize some gesture in the future.

The paper is organized as follows: second chapter introduces related works and the outline of our method; the third chapter describes our method in detailed; the fourth chapter includes experiments; we give the conclusion of our method in the last chapter.

II. RELATED WORK

Our work is based on hand edge detection. Relevant work to this subject is what has been done for edge detection and skin detection. There are some main researches in these fields.

Edge detectors are algorithms which aim to identify points in a digital image where the image brightness changes sharply, or more formally has discontinuities. Discontinuities in image brightness are likely to correspond to: discontinuities in depth, discontinuities in surface orientation, and changes in material properties and variations in scene illumination.

There are many methods for edge detection, but most of them can be grouped into two categories, search-based and zero-crossing based.

The search-based methods detect edges by first computing a measure of edge strength, usually a first-order derivative expression such as the gradient magnitude, and then searching for local directional maxima of the gradient magnitude using a computed estimate of the local orientation of the edge, usually the gradient direction.

The zero-crossing based methods search for zero crossings in a second-order derivative expression computed from the image in order to find edges, usually the zero-crossings of the Laplacian or the zero-crossings of a non-linear differential expression.

The Canny edge detector is the most widely used algorithm for edge detection. It is known as the best edge detector and it presents a multiscale approach. Using canny detector the obtained contours are thinner. In [7], is a list of criteria that Canny used to improve the edge detection. The first one is low error rate, which means real edges must be detected while no responses should occur to non-edges. The second criterion is that the edge points must be well localized,

[1] Facultatea de Electronică şi Telecomunicaţii, Departamentul Comunicaţii Bd. V. Pârvan Nr. 2, 300223 Timişoara, e-mail georgiana.simion@etc.utt.ro
[2] Facultatea de Electronică şi Telecomunicaţii, Departamentul Comunicaţii Bd. V. Pârvan Nr. 2, 300223 Timişoara, e-mail vasile.gui@etc.utt.ro
[3] Facultatea de Electronică şi Telecomunicaţii, Departamentul Comunicaţii Bd. V. Pârvan Nr. 2, 300223 Timişoara, e-mail marius.otesteanu@etc.utt.ro
[4] Facultatea de Electronică şi Telecomunicaţii, Departamentul Comunicaţii Bd. V. Pârvan Nr. 2, 300223 Timişoara, e-mail gheorghe.popa@etc.utt.ro
[5] Facultatea de Electronică şi Telecomunicaţii, Departamentul Comunicaţii Bd. V. Pârvan Nr. 2, 300223 Timişoara, e-mail ciprian.david@etc.utt.ro

the distance between the edge pixels as found by the detector and the actual edge has to be minimum. A third criterion involves one response to a single edge, in order to have this canny edge detector first smoothes the image to eliminate and noise; then finds the image gradient to highlight regions with high spatial derivatives.

Skin color has proven to be a useful cue for hand/face detection, localization and tracking. Numerous colorspaces for skin modeling have been proposed, ones of the most popular color spaces are: RGB, HIS, HSV, HSL, TLS, YCrCb.

RGB is a colorspace originated from CRT display applications, when it was convenient to describe color as a combination of three colored rays (red, green and blue). It is one of the most widely used colorspaces for processing and storing of digital image data.

Hue-saturation based colorspaces were introduced when there was a need for the user to specify color properties numerically. They describe color with intuitive values, based on the artist's idea of tint, saturation and tone.

Several interesting properties of Hue were noted in [8]: it is invariant to highlights at white light sources, and also, for matte surfaces, to ambient light and surface orientation relative to the light source.

$$H = \arccos \frac{\frac{1}{2}((R-G)+(R-B))}{\sqrt{((R-G)^2+(R-B)(G-B)}}$$

$$S = 1 - 3\frac{\min(R,G,B)}{R+G+B} \qquad (1, 2, 3)$$

$$V = \frac{1}{3}(R+G+B)$$

A normalized chrominance-luminance TSL space is a transformation of the normalized RGB into more intuitive values, close to hue and saturation in their meaning.

$$S = [9/5(r'^2 + g'^2)]^{1/2} \qquad (4)$$

$$T = \begin{cases} \arctan(r'/g')/2\pi + 1/4, g' > 0 \\ \arctan(r'/g')/2\pi + 3/4, g' < 0 \\ 0, g' = 0 \end{cases} \qquad (5)$$

$$L = 0.299R + 0.587G + 0.114B \quad \text{where} \qquad (6)$$

$$r' = r - 1/3$$
$$g' = g - 1/3 \qquad (7, 8)$$

$$r = \frac{R}{R+G+B}; g = \frac{G}{R+G+B} \qquad (9,10)$$

RGB, HS and Hue colour spaces have shown the best results for colour skin segmentation.

YCrCb is an encoded nonlinear RGB signal, commonly used by European television studios and for image compression work. Color is represented by luma (which is luminance, computed from nonlinear RGB [9]), constructed as a weighted sum of the RGB values, and two color difference values Cr and Cb that are formed by subtracting luma from RGB red and blue components.

$$Y = 0.299R + 0.587G + 0.114B$$
$$C_r = R - Y \qquad (11, 12, 13)$$
$$C_b = B - Y$$

The main goal of a skin color detector is to produce a rule in order to discriminate between skin and non-skin. There are several methods [10] to model the skin. One method to build a skin classifier is to define explicitly the boundaries skin cluster in some colorspace.

The non-parametric skin modelling methods key idea is to estimate skin color distribution from the training data without deriving an explicit model of the skin color. The result of these methods sometimes is referred to as construction of Skin Probability Map (SPM) [11], [12] - assigning a probability value to each point of a discretized colorspace.

The advantages of the non-parametric methods are : they are fast in training and usage and they are theoretically independent to the shape of skin distribution (which is not true for explicit skin cluster definition and parametric skin modeling). The disadvantages are: much storage space required and inability to interpolate or generalize the training data.

The need for more compact skin model representation for certain applications along with ability to generalize and interpolate the training data stimulates the development of parametric skin distribution models.

Among the parametric methods are: single Gaussian, mixture of Gaussians, multiple Gaussian clusters and elliptic boundary model.

Dynamic skin distribution models are fast in both training and classification and there are able to update to changing conditions.

There are several papers like [13, 14, 15] in which gestures are classify using the edges.

In [13] the hand is located in the image and segmented from the background before recognition, and then a Gaussian model is chosen to represent the skin-colour probability density function. The blob representation of the hand is obtained applying a connected components algorithm to the probability image, which groups pixels into the same blob.

In [14] after the hand has been segmented, a model based approach based on the Hausdorff distance that works on edge map images and a visual memory is used to recognize the hand posture.

In [15] hand contour is used as image feature for recognition of hand posture, the Fourier Transform is apply to all of points along the contour, and use Fourier Descriptors (FDs) as feature vector of hand posture.

III.OUR WORK

The main idea of this work is to exploit the knowledge of skin detection techniques in order to improve the performance of hand contour detection. A specialized technique to detect hand edges in a particular context was designed. The resulting technique is not a general skin detector algorithm; it was developed for our specific application. It was designed to work optimally with several backgrounds. The hand's backgrounds considered are not very variegated but not very simple either.
Modelling skin colour requires the selection of an appropriate colour space and identifying the cluster associated with skin colour in this space. HSV space (Hue, Saturation and Value) was chosen because hue is less sensitive to different skin colors and because it is more robust to illumination changes. It was observed [16] that the hue of human skin is the same for all the races, except the albinos. Thus, colours can be specified using just two parameters instead of the three specified by RGB space colour (Red, Green, Blue).
In order to define the skin model, several samples of skin were used to get a skin collage. The H range found was [0÷0.1; 0.9÷1] and S range found was [0.1÷ 0.6]. The V values depended a lot on the illumination level. Because we wanted a model independent of the illumination level, the V value was eliminated. Although in some papers only H values were used for treckers, we founded based on our hands database experiments the S range [0.1÷ 0.6] and we wanted to exploit this information. The colour space that we used was HS.

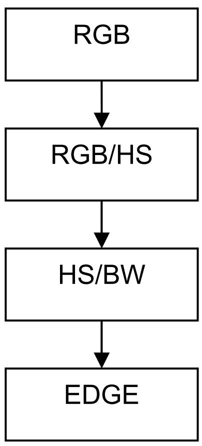


Fig .1 Framework diagram

Based on the skin model, a framework (see fig.1) was developed to detect edges in image sequences. The main problem was to find an optimal mapping from HS images to the BW space in order to make possible the use of the monochromatic canny edge detector.
First the RGB images were mapped in HS space; second, an optimized HS to BW space transformation was searched, with the objective of maximizing correct hand contour detection. We adopted a polynomial transform model, of the form:

$$\alpha *h*h + \beta *s*s + \gamma *(h*h+s*s)*(h*h+s*s); \qquad (14)$$

with parameters α, β, γ. Polynomial fitting is commonly used in fitting.
We were looking for an optimized set of parameters, maximizing the average contrast between skin and nonskin areas. The contrast is defined as:

$$dif=abs(2*(t1-t2))/sqrt(\alpha * \alpha + \beta * \beta + \gamma * \gamma); \qquad (15)$$

where:
$$t1=\alpha*Ha*Ha+\beta*Sa*Sa+\gamma*(Ha*Ha+Sa*Sa)*(Ha*Ha+ + Sa*Sa); \qquad (16)$$
represents the skin model , Ha and Sa are medium hue /saturation value for skin. The denominator was introduced to avoid unlimited global amplitude grow.

$$t2= \alpha *h*h + \beta *s*s + \gamma *(h*h+s*s)*(h*h+s*s); \quad (17)$$
is the current pixel which is analyzed.
In order to find the set of parameters maximizing (15), several optimization algorithms can be used. Optimization algorithms have the following three subcategories: combinatorial optimization, dynamic programming, and evolutionary algorithms. We used a genetic algorithm because of their ability to avoid being trapped in local extremes of the search space. As a general rule of thumb genetic algorithm might be useful in problem domains that have a complex fitness landscape where simpler optimization algorithm such as hill climbing algorithms has increased chances to fail to detect the global optimal solution. Genetic algorithms are a particular class of evolutionary algorithms that use techniques inspired by evolutionary biology [17, 18, 19] such as inheritance, mutation, selection, and crossover or recombination. A typical genetic algorithm requires two things to be defined: a genetic representation of the solution domain and a fitness function to evaluate the solution domain.
The fitness function which has been used is:
$$ftnsg = \Sigma \exp(-dif); \qquad (18)$$
where dif was defined in (15) .

V. EXPERIMENTS

In our experiments several backgrounds have been used and there are shown in Fig. 2. All these backgrounds are neither very complex but nor simple. As it can be seen in these backgrounds there are darker colours and lighter colours.The parameters α, β, γ have been found running the genetic algorithm with fitness function defined in (18) and 3 independent variables for the fitness function. Population type was defined as double vector, initial population size was 20 and the crossover fraction 0.8000, selection - uniform. With α, β, γ parameters known, the mapping from HS to BW has been done.

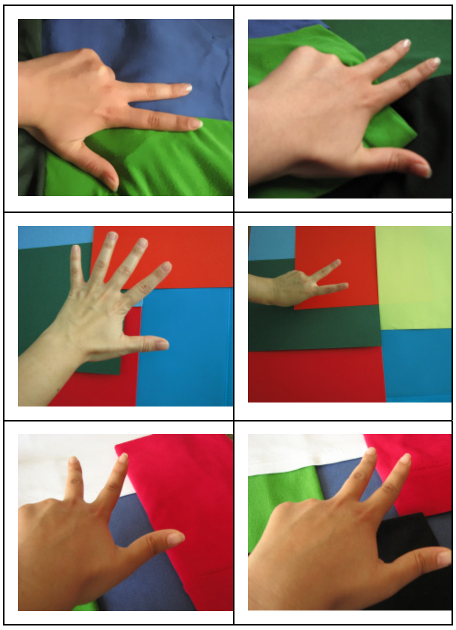


Fig.2 Examples of backgrounds

In Fig. 3 a) can be seen the RGB input image and in b) there is the HS image mapped in BW space with our specific transformation, e) represents the RGB input image mapped in BW without using our transformation, c)and f) are the edges detected on both HS mapped in BW and BW images. The polynomial transform model's parameters are α=0.4881, β=0.6562, γ=0.2873.The canny edge detector has been used to obtain thin hand edges.

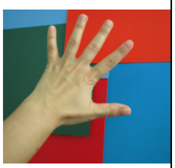
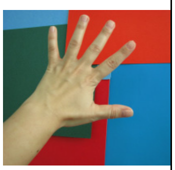
		
a) RGB image	b)HS image mapped in BW space	c)Canny=0.35
		
d)RGB image	e)Gray scale image	f)Canny=0.35

Fig. 3. The polynomial parameters are α=0.4881, β=0.6562, γ=0.2873

In Fig. 4 can be seen the contours' hands for the HS images mapped in BW and in Fig. 5 are shown the contours for the same images, the BW images. The hands edges have been better extracted with our specific transformation; less false edges have been found on hands. The backgrounds have more false contours because the transformation amplifies the local contrast. These edges can be ignored to the semantic level or with a background technique.

VI.CONCLUSIONS

In this paper a method designed to optimize hand contour detection is presented. The main problem was to find an optimal mapping from HS images to the BW space to make possible the use of the monochromatic canny edge detector. In order to do this we adopted a polynomial transform model and a set of parameters were looking to maximize the average contrast between skin and non-skin areas The contours hands have been better detected and false hands edges have been removed. Some false background edge which can be removed with a compositional approach and/or background techniques occurred.

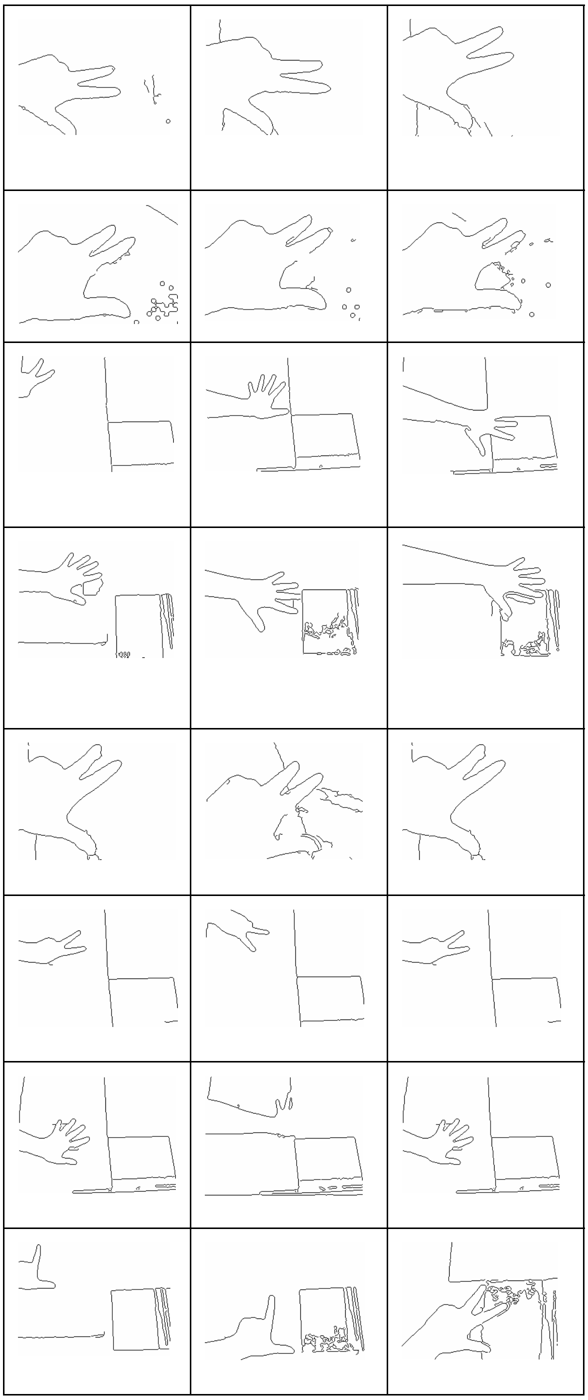


Fig. 4. Hands contours for the HS images mapped in BW

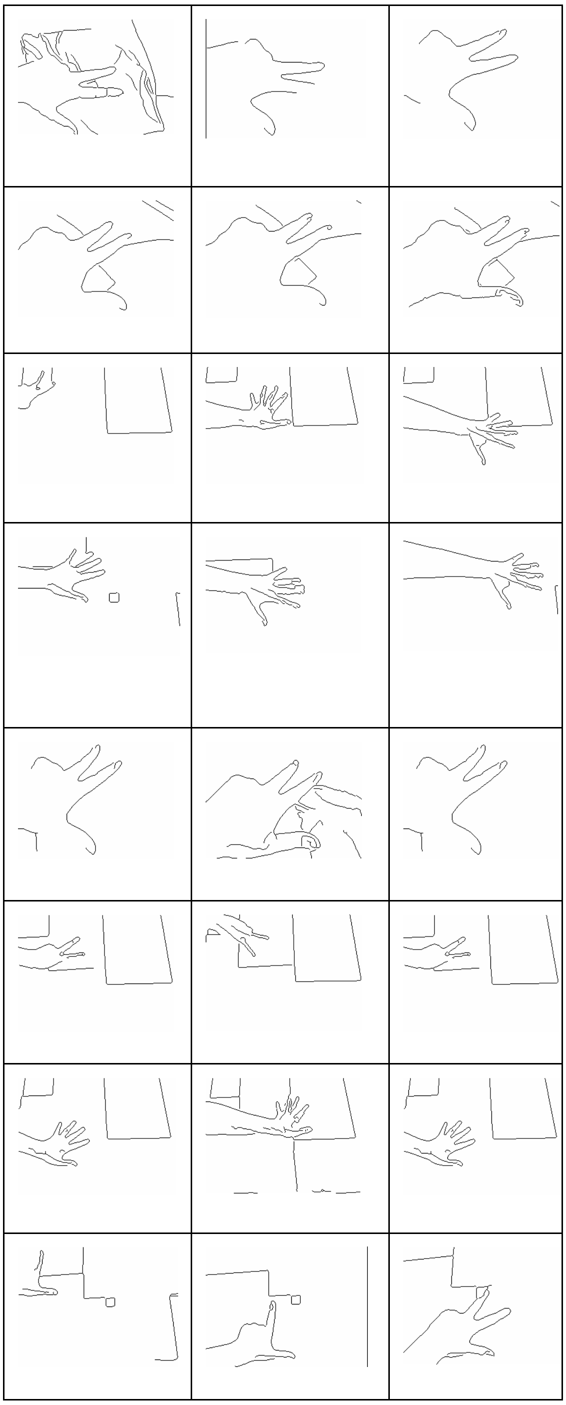


Fig. 5 Hands contours for the gray scale images.

ACKNOLEDGMENT

This study was conducted in the framework of the research contract for young Ph.D. students, no.185/2007 sponsored by CNCSIS.

REFERENCES

[1] R. Fergus1 L. Fei-Fei P. Perona A. Zisserman"Learning Object Categories from Google's Image Search "Proceedings of the Tenth IEEE International Conference on Computer Vision - Volume 2, Pages: 1816 – 1823, 2005 ISBN ~ ISSN: 1550-5499 , 0-7695-2334-X-02

[2] Bjorn Ommer Joachim M. Buhmann "Learning the Compositional Nature of Visual" CVPR'07, IEEE, 2007.

[3] Bjorn Ommer Joachim M. Buhmann" Learning Compositional Categorization Models Objects" ECCV'06, LNCS 3953, Springer, 2006

[4]Li Fei-Fei, Pietro Perona "A Bayesian Hierarchical Model for Learning Natural Scene Categories "EEE Comp. Vis. Patt. Recog. 2005.

[5] Antonios Oikonomopoulos, Ioannis Patras, Maja Pantic "Spatiotemporal Salient Points for Visual Recognition of Human Actions", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 36, NO. 3, JUNE 2006

[6] Ivan Laptev, Tony Lindeberg" Space-time Interest Points " Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV'03),0-7695-1950-4/03

[7] J Canny "A computational approach to edge detection" IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 8 , Issue 6 (November 1986)

[8] Skarbek, W., and Koschan, A. 1994. Colour image segmentation – a survey . Tech. rep., Institute for Technical Informatics, Technical University of Berlin, October.

[9] Poynton, C. A. 1995. Frequently asked questions about colour. ftp://www.inforamp.net/pub/users/ poynton/ doc/colour/ColorFAQ.ps.gz.

[10] Vladimir Vezhnevets Vassili Sazonov Alla Andreeva" A Survey on Pixel-Based Skin Color Detection Techniques" Machine Graphics & Vision International Journal archive Volume 14 , Issue 1 (January 2005),Pages: 61 - 70 ISSN:1230-0535

[11] Gomez, G., Morales, E.. Automatic feature construction and a simple rule induction algorithm for skin detection. In Proc. of the ICML Workshop on Machine Learning in Computer Vision, 31–38. 2002

[12] Gomez, G. On selecting colour components for skin detection. In Proc. of the ICPR, vol. 2, 961–964. 2000.

[13] Cristina Manresa, Javier Varona, Ramon Mas and Francisco J. Perales"Hand Tracking and Gesture Recognition for Human-Computer Interaction" Electronic Letters on Computer Vision and Image Analysis 5(3):96-104, 2005

[14] Farhad Dadgostar, Abdolhossein Sarrafzadeh "A Component-based Architecture for Vision-based Gesture Recognition" IVCNZ 2005, pp. 322-328 Dunedin, New Zealand.

[15] Elena Sánchez-Nielsen, Luis Antón-Canalís, Mario Hernández-Tejera" Hand Gesture Recognition for Human-Machine Interaction"

[16] G. R. Bradski," Computer vision face tracking as a component of a perceptual user interface", Intel Technology Journal Q2 1998, http://developer.intel.com/technology/itj/archive/1998.htm.

[17] Goldberg, David E., Genetic Algorithms in Search, Optimization & Machine Learning, Addison-Wesley,1989.
[18] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. "A Globally Convergent Augmented Lagrangian Algorithm for Optimization with General Constraints and Simple Bounds", SIAM Journal on Numerical Analysis, Volume 28, Number 2, pages 545–572,1991.
[19] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. "A Globally Convergent Augmented Lagrangian Barrier Algorithm for Optimization with General Inequality Constraints and Simple Bounds", Mathematics of Computation, Volume 66, Number 217, pages 261–288, 1997.