archive ouverte UNIGE
http://archive-ouverte.unige.ch

Article

Tactile artificial sight : segmentation of images for scene simplification

PUN, Thierry

Abstract

Aiming at designing the image processing unit of a visual prosthesis for sight handicapped, an efficient picture simplification scheme for tactile outputs is proposed. Some psychological considerations are given to help in its development. A new simple and robust segmentation method which combines the advantages of both the classical edge and region approaches is introduced. This method uses, on the one hand, the inherent property of edges to provide transition thresholds between regions of monotonous intensity, and on the other hand, the fact that region segmentation methods give closed regions. A region labeling is applied using those thresholds, yielding well outlined areas. Artificial textures are introduced to help in the tactile discrimination of shapes.

Reference

PUN, Thierry. Tactile artificial sight : segmentation of images for scene simplification. *IEEE transactions on biomedical engineering*, 1982, vol. BME 29, no. 4, p. 293-299

DOI : 10.1109/TBME.1982.324947

Available at:
http://archive-ouverte.unige.ch/unige:47455

Disclaimer: layout of this document may differ from the published version.

UNIVERSITÉ DE GENÈVE

Tactile Artificial Sight: Segmentation of Images for Scene Simplification

THIERRY PUN, STUDENT MEMBER, IEEE

Abstract—**Aiming at designing the image processing unit of a visual prosthesis for sight handicapped, an efficient picture simplification scheme for tactile outputs is proposed. Some psychological considerations are given to help in its development. A new simple and robust segmentation method which combines the advantages of both the classical edge and region approaches is introduced. This method uses, on the one hand, the inherent property of edges to provide transition thresholds between regions of monotonous intensity, and on the other hand, the fact that region segmentation methods give closed regions. A region labeling is applied using those thresholds, yielding well outlined areas. Artificial textures are introduced to help in the tactile discrimination of shapes.**

I. Visual Prosthesis Research

A. Background

The field of sensory prosthesis for sight handicapped has been tremendously explored for a long time [1], [2]. Despite a lot of research, the problem of "showing" the blind an image has, however, still no answer. Most of the already existing devices [3] have been designed for specific purposes and they replace a very restricted part of vision, such as mobility aids [4] (laser cane, ultrasonic binaural sensor . . .) and reading aids [3] (Optacon text reader, Kurzweil speaking machine . . .). Current investigations actually stress complex prosthetic systems, like neurological stimulators of the visual cortex [5] and tactile vision systems [6], [7]. In addition to technological difficulties, the development of these systems raises several psychological problems. This work tries to incorporate such considerations to help in the development of a tactile prosthesis.

Tactile perception of simple patterns is a well-known information acquisition mean for the sight handicapped. The Braille alphabet and relief maps of cities are some examples. Experiments with the tactile perception of more complex images have been performed some years ago by Bach-Y-Rita and others [6] with a rather mitigated success. Such tactile prosthesis are composed of an image sensor, a processing unit, and a tactile display [8]. This paper describes the conception of algorithms for the image processing unit.

B. Psychological Considerations and Adequate Image Processing

From psychologists, the partial success of Bach-Y-Rita's first experimentations was probably partly due to the lack of an evaluation of the educational psychology problems of handicapped. Complex tactile images could be made perceptible to a blind person only by giving him access to the relation between what he does and what he feels [9]. Practically speaking, this may signify that the subject will be able to operate the image sensor himself. He will extract the transformation law between his action on the camera and the evolutive tactile pattern he feels; this *perceptual invariants* extraction will enable the tactile picture understanding process. However, this *evolutive* prosthesis would imply a rather formidable learning task.

The work presented here aims at a different application: the development of an image processor yielding tactile images for educational purposes. Such a machine would be used in classes for blind children, providing relief illustrations from grey-level documents. This will help the children to develop their mental images, facilitating experiments and coordination of movements. From concerned teachers there is a real need of such an "on-line" portable device. Relief printers, yielding 0.2 to 0.6 mm high outputs, are close to being available [10].

For the above mentioned *evolutive* approach, there is perhaps no direct need of complex image processing, the subject's sensory system being solicitated to do the most of the task. For our application, however, the tactile image has to be simplified: this is the *Gestalt* approach. Preliminary tests with blind subjects show that the pictorial information has to be processed in order to remove irrelevant details while emphasizing informative patterns like the contours of objects. These contours have to be closed to delimite and separate semantic entities. Synthetic textures have also to be introduced to help in the tactile discrimination of shapes. In order to obtain closed contours of informative areas, a new edge/region method for image segmentation has been designed. It avoids iterative processes, therefore it is more convenient for a simple portable implementation than more efficient but more complex algorithms.

II. A New Edge/Region Method for Image Segmentation

A. The Segmentation Problem

Image segmentation is the division of an image into different regions, each one being characterized by the uniformity of certain features [11]. Only the intensity feature (grey-level) is considered in our context. Segmentation research [12] has concentrated on two approaches: edge and region algorithms. The edge approach first extracts picture elements (pixels) of high gradient values, then thresholds out the weak ones to get contours. However, they are usually broken and their reconstruction is extremely difficult.

The duals of this approach are region growing and region

Manuscript received November 30, 1981. This work was supported by the Swiss National Research Foundation under Grand 3.883.079.

The author is with the Signal Processing Laboratory, Swiss Federal Institute of Technology, Lausanne, Switzerland.

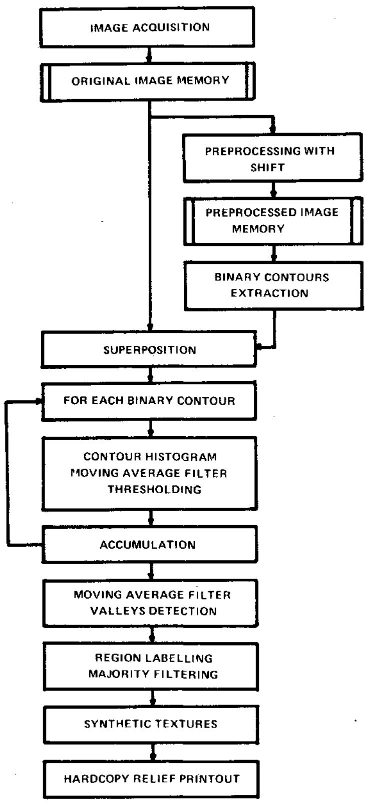


Fig. 1. Block diagram of the image processing unit.

labeling. Region growing starts at some pixel of well-defined feature value(s) and creates regions by merging adjacent points satisfying an uniformity criterion. Such are split-and-merge algorithms. They are generally quite time and memory consuming. These algorithms are inherently well suited for relaxation procedures. Region labeling, also called global feature thresholding, consists in identifying thresholds on (in our case) the grey-level histogram, usually at valleys between the modes. The picture is then scanned, and a label is attributed to each point depending to which slice of grey-levels it belongs. This method, as region growing, yields closed regions (which at the limit may be one point wide). It is, however, usually difficult to determine the thresholds, because the modes overlap or are hidden by noise. There is then a need of modification of this histogram in order to induce or to find modes [13].

B. The Segmentation Method

The approach presented here combines advantages of both edge extraction and region labeling methods. The inherent property of edges to easily provide transitions between regions of uniform intensity enables the determination of thresholds corresponding to these intensities. A region labeling is then applied with those thresholds, yielding closed regions. Neither modification nor noise deconvolution of the grey-level histogram is requested to obtain the thresholds. Although only the grey-level is used as feature, the method can easily be extended to multifeature segmentation.

Instead of using the normal grey-level histogram which may be problematic to process, only some pixels inside the regions of constant feature are used. These points can be found just beside the edges. The method consists in extracting skeletonized contours,[1] then to shift them with respect to the edges. Local contour intensity histograms are computed on the original image *at addresses pointed by these skeletonized contours*; due to the way they are determined they are usually unimodal, sometimes bimodal. These contour histograms (one for each contour) are thresholded and accumulated, providing a density function where modes are quite separated. Segmentation thresholds are determined at the valleys, and a staircase look-up function (in fact a requantization law) is built, each step corresponding to one of the thresholds.

The shift between the edge image and the contour image is done when preprocessing the picture. If the resulting value of the convolving operator is attributed to the upper left-hand corner of the window, each convolution shifts the filtered image by the half of the mask size. Fig. 1 presents the block diagram of the method.

III. Implementation and Results

A. Preprocessing

One of the test images used for this study is the cameraman [Fig. 2(a)]. It is digitized with a raster of 256×256 points and quantized to 8 bits. Its grey-level histogram is shown in Fig. 2(b). It is not a typically relevant test image in the context of our work because it introduces quite complex spatial notions. However, its complexity makes it suitable to evaluate the goodness of the methods. The algorithms yielding the skeletonized contours [14], [15] are briefly recalled here.

First, median smoothing [16] is applied to reduce the high frequency noise without blurring edges. The size of the filter is a function of the details that have to be kept or removed. It cannot be too large because this would distord the shapes, but if too small it would have no effect. This size has been experimentally fixed to 5×5 points, providing a shift of two pixels.

The edge extraction is then performed by applying the usual 2×2 high-pass Mero–Vassy operator. It replaces the level $z(i, j)$ at point of coordinates (i, j) by

$$\begin{aligned} z'(i, j) = &|z(i, j) + z(i, j+1) - z(i+1, j) - z(i+1, j+1)| \\ &+ |z(i, j) + z(i+1, j) - z(i, j+1) \\ &- z(i+1, j+1)|. \end{aligned} \tag{1}$$

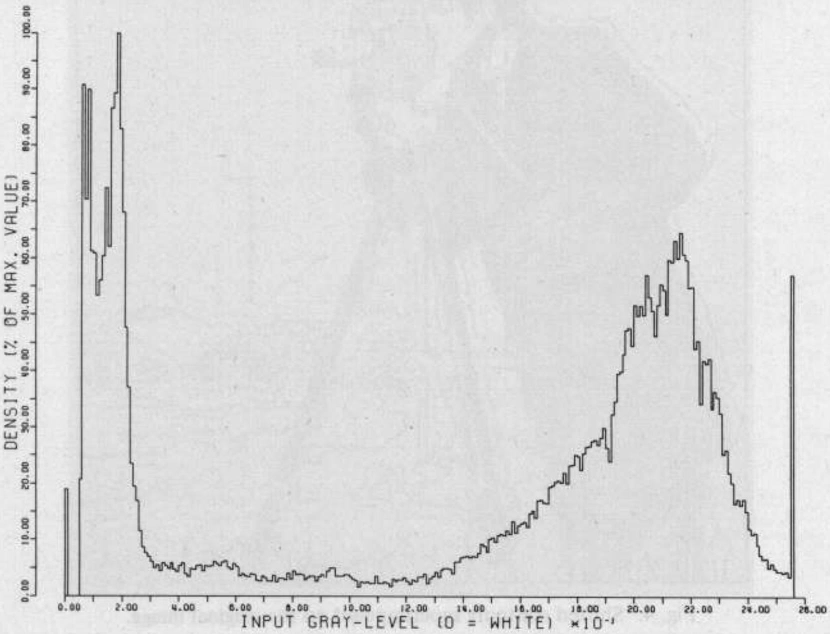
The shift is there one point, the total shift of the whole preprocessing being three points.

B. Contours Extraction

The ridge-riding algorithm [17] provides directly skeletonized contours. It is applied on the preprocessed image, which is seen as a surface where the ridges are the edges. The procedure requires two thresholds: one above which a track begins, following a ridge, and the other below which the track stops. This algorithm yields skeletons which are probably the most significant ones, that is the ridges of the edges. It is robust because it detects contours even in very poorly contrasted

[1] In the text edges will refer to the high-gradient points, while contours will refer to the binary skeletonized edges.

(a)



(b)

Fig. 2. The cameraman. (a) Original grey-level image. (b) Grey-level histogram.

areas. Fig. 3 shows the result of the algorithm, after elimination of contours shorter than 30 points.

The two thresholds are selected via the entropic thresholding method [14], which is very easy to implement. Let us consider the grey-level histogram of the preprocessed image. The probability of occurrence of each of its $n + 1$ levels is denoted by pi. Find the two levels $na1$ and $na2$ dividing the histogram into three parts containing the same number of points ($na1$ and $na2$ are the $\frac{1}{3}$-fractiles). Then two anisotropy coefficients $a1$ and $a2$ are defined by

$$a1 = -\sum_{0}^{na1} (pi \cdot lbpi)/H \qquad a2 = -\sum_{na1+1}^{na2} (pi \cdot lbpi)/H \qquad (2)$$

where $lbpi$ is the logarithm in base 2 of pi, and H the Shannon

Fig. 3. Binary contours longer than 30 points (after the ridge riding).

Fig. 4. Shifted contours superimposed on the original image.

entropy of the histogram

$$H = -\sum_{0}^{n} pi \cdot lbpi. \tag{3}$$

Note that $-lbpi$ can be approximed by the number of bits of $1/pi$. The thresholds $t1$ and $t2$ are selected as the levels satisfying

$$\sum_{0}^{t1} pi = \frac{1}{3} + |\frac{1}{3} - a1| \quad \sum_{0}^{t2} pi = \frac{2}{3} + |\frac{1}{3} - a2|. \tag{4}$$

If the histogram has three equally filled modes, the thresholds are selected in the two valleys. The more asymmetric it becomes, the more shifted toward the high peaks they are.

C. Segmentation

Fig. 4 shows the contours superimposed on the original image. One histogram is computed for each of them. These histograms are smoothed with a moving average (MA) filter whose size L is proportional to their entropy H. This yields a small window when only few levels are present, avoiding spurious blurring, while in case of many levels a large L results in a good smoothing. The ratio between L and H has been experimentally fixed to 2 (L is $2H$). This choice provides a

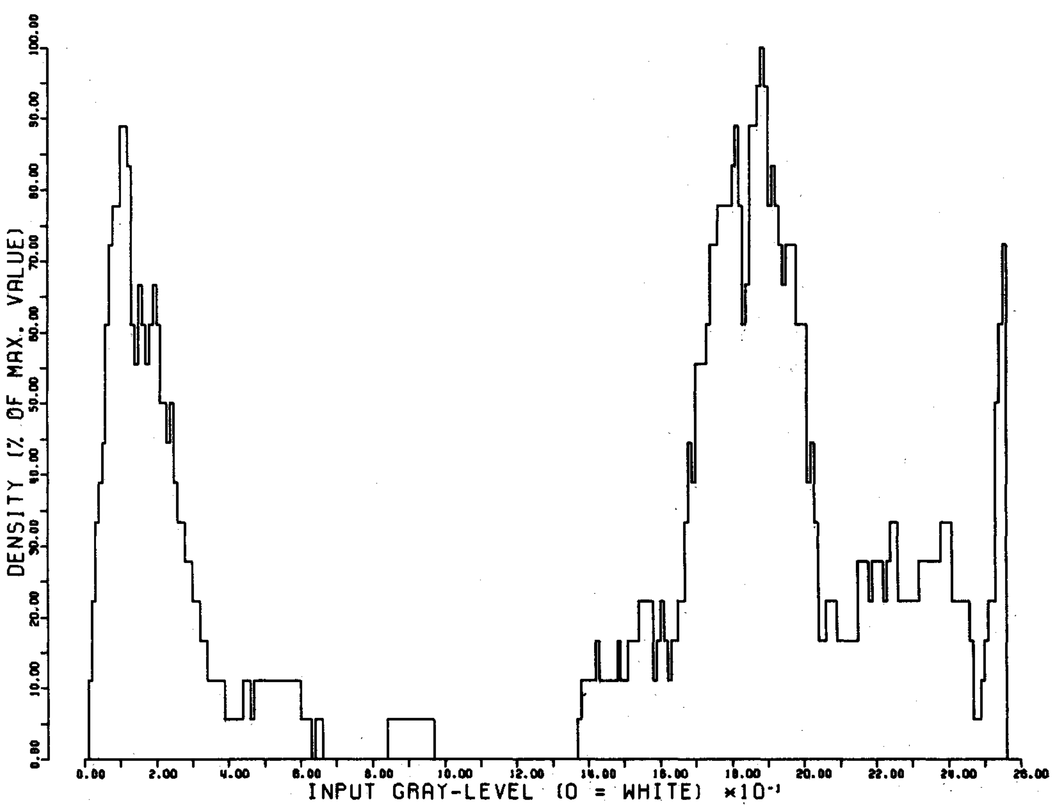


Fig. 5. Accumulation of the binary contour histograms.

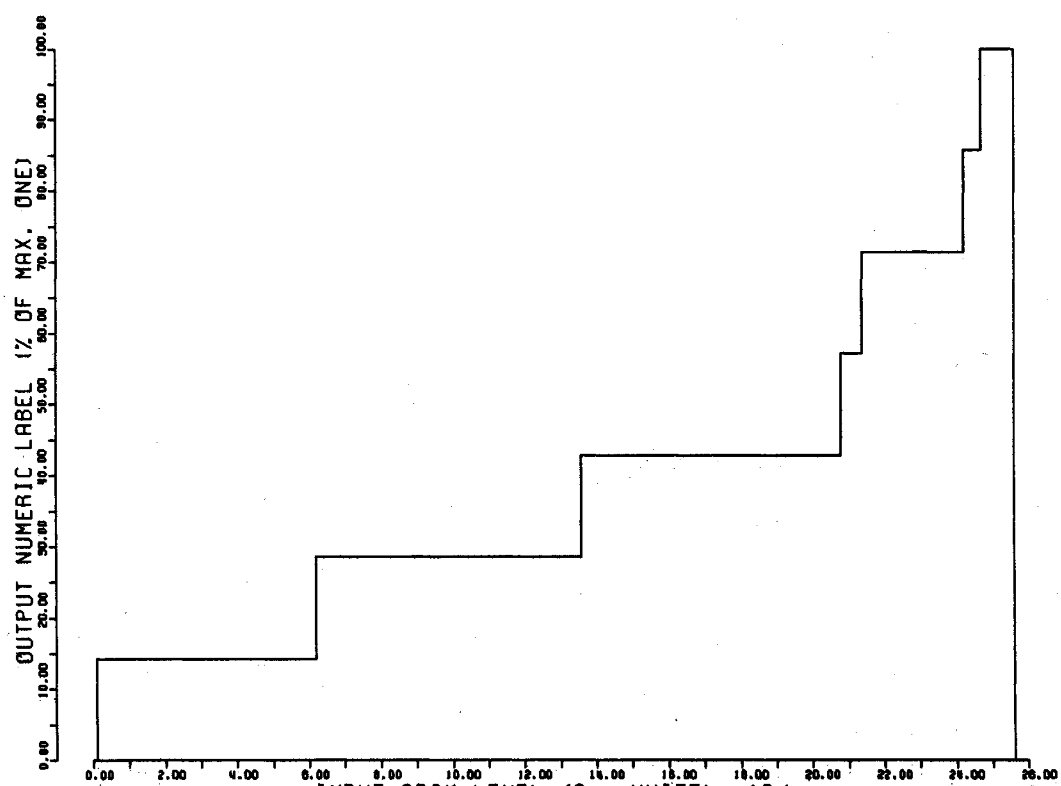


Fig. 6. Look-up function.

good compromise between a too strong smoothing (greater ratios), which alters the selection of thresholds and the acceptance of spurious noise.

After MA filtering, each histogram is thresholded in order to detect the main(s) mode(s). The threshold is fixed at half of the maximal value of each of them, but the results are not sensitive to this choice. All these binary contour histograms are finally summed (Fig. 5). The MA smoothing is then again applied. Thresholds for the segmentation are determined as the valleys (if they are large, each of their extremity is accepted). The staircase look-up function is then built, each step corresponding to one of the thresholds (Fig. 6).

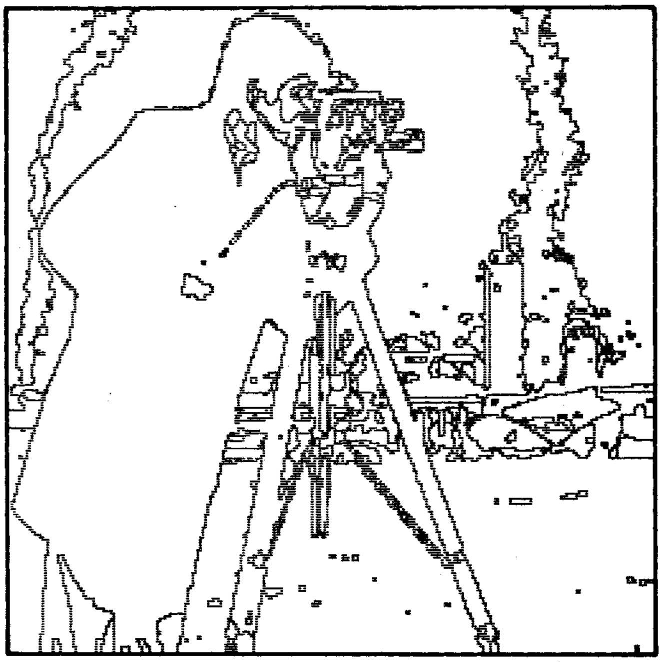


Fig. 7. Borders of the labeled regions after a 3 × 3 majority filtering.

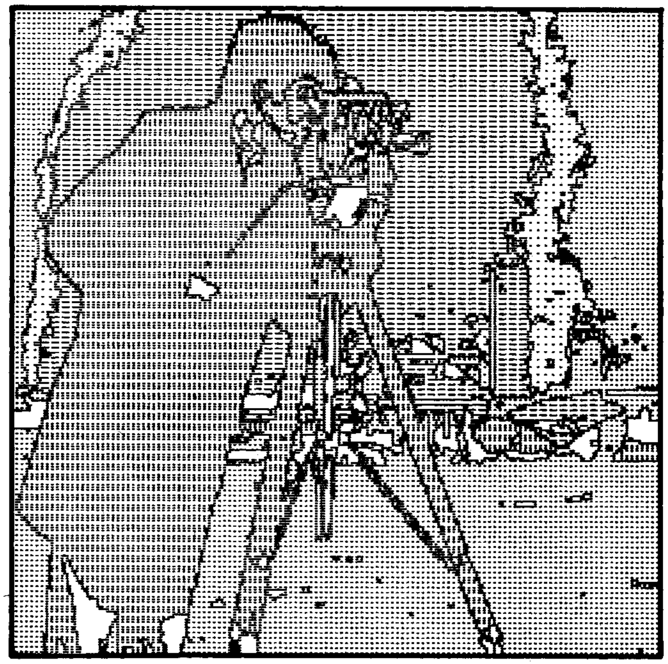


Fig. 8. Introduction of synthetic textures (four different patterns).

After labeling, a post-processing majority filter is applied to remove the very small regions and to provide large areas which are more easily tactily interpretable. The label at the center of a convolving window is replaced by the most occurring one inside this neighborhood. The extent of this window has been fixed to 3 × 3 points (larger sizes distord too much shapes). Since this majority filtering converges quickly, it has been only applied once. Fig. 7 shows the borders between the regions of same label.

IV. Results and Concluding Comments

A. Discussion and Comments

This segmentation method leads to good results. Poorly contrasted objects, such as the two towers in the right part of the picture, are segmented. The general shape of the elements is well outlined. There are, however, two kinds of errors. First, semantical ones which are not due to the segmentation itself. This is the case for the lines rising up in the sky. The regions they border have the same levels as the towers. It is then impossible to segment these buildings without dividing the sky into distinct parts. Second, there are errors due to the implementation of the method. For example, dark parts such as the man's pocket have disappeared because there was no contour extracted along it (Fig. 3). There was thus no contour histogram associated with it. Such an error could be avoided by tolerating contours shorter than 30 points. This was not suitable in our context, with too much details appearing then.

For elementary pictures it is possible to detect the valleys directly on the grey-level histogram. But for more complex images, this method fails. The proposed method provides histograms with modes more separated than on the original ones (compare Figs. 2(a) and 5). More complicated methods [11] probably yield better results; however, the proposed segmentation is simple, robust, and provides convenient pictures for our purpose. It also could be easily extended to multifeature segmentation.

As mentioned above, synthetic textures have to be created to ease in the tactile discrimination of regions. It has been chosen to texture only the largest ones (supposed to be those with the most occurring labels), which are difficult to be tactily discriminated. Fig. 8 shows the picture 7 filled in with four different synthetic patterns.

This method has been experimented on three complex images and on several single pictures (animals, objects . . .), always with good segmentation. It is actually implemented on a CDC Cyber 170-720 computer; however, a portable realization is planned for the near future.

B. Conclusion

Aiming at designing the image processing unit of a visual prosthesis for sight handicapped, an efficient picture simplification scheme for tactile outputs is proposed. A new, simple, fast, and robust segmentation method which combines the advantages of both the classical edge and region approaches is introduced. This method uses, on the one hand, the inherent property of edges to provide transition thresholds between regions of monotonous intensity, and on the other hand, the fact that region segmentation methods give closed regions. A region labeling is applied using those thresholds, yielding well-outlined areas. Artificial textures are introduced to help in the tactile discrimination of the shapes.

Acknowledgment

The author wishes to thank Prof. F. de Coulon, Prof. M. Kunt, and Prof. A. Bullinger from the Faculty of Psychology in Geneva, Switzerland, and M. Ekelof from the Center for Sight Handicapped in Lausanne, Swizerland, for numerous helps all along this work. Thanks also to A. Ligtenberg for help in preparing this paper.

REFERENCES

[1] Gallois, "Anaculoscope de C. Grim," *Bull. le Valentin Hauy*, pp. 76–78, Oct. 1883 (Note: probably the first electromechanical tactile visual prosthesis that has been designed!).

[2] E. E. Fournier d'Albe, "The Optophone: An instrument for reading by ear," *Nature*, vol. 105, May 1920.

[3] J. A. Brabyn, "Artificial vision: A review of sensory aids for the blind," in *Proc. IEEE 1980 Frontiers in Eng. in Health Care, 2nd Annu. Conf. of Eng. in Med. Biol. Soc.*, Sept. 1980, pp. 98–104.

[4] H. Freiberger, "Mobility," in *Proc. Symp. on Prosthetic Aids for Blind*; also in *Trans. Amer. Acad. Opht. Otol.*, vol. 78, Sept. 1974.

[5] W. H. Dobelle *et al.*, "Artificial vision for the blind by electrical stimulation of the visual cortex," *Neurosurgery*, vol. 5, no. 4, 1979.

[6] P. Bach-Y-Rita, "Visual information through the skin," in *Proc. Symp. Prosthetic Aids for Blind*, Sept. 1980.

[7] T. Pun, "Visual prosthesis: Digital image processing for tactile conversion," *Sensory World*, no. 41, Dec. 1980.

[8] "Dynamic tactile displays in man-machine systems," Special Issue, *IEEE Trans. Man-Mach. Syst.*, vol. MMS-11, Mar. 1970.

[9] G. Jansson, "Human locomotion guided by a matrix of tactile point stimuli," in *Active Touch*. New York: Pergamon, 1978, pp. 263–271.

[10] R. Erhardt *et al.*, "Fast and economic conversion of images to a tactile presentation for blind people," *Med. Prog. Technol.*, vol. 6, pp. 123–130, 1979.

[11] E. M. Riseman and M. A. Arbib, "Survey: Computational techniques in the visual segmentation of static scenes," *Comput. Graphics Image Proc.*, vol. 6, pp. 221–276, 1977.

[12] K. S. Fu and J. K. Mui, "A survey on image segmentation," *Pattern Recog.*, vol. 13, pp. 3–16, Jan. 1981.

[13] J. S. Weszka and A. Rosenfeld, "Histogram modification for threshold selection," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, Jan. 1979.

[14] T. Pun, "Entropic thresholding, A new approach," *Comput. Graphics Image Proc.*, vol. 15, pp. 210–239, 1981.

[15] T. Pun and F. de Coulon, "Image processing for visual prosthesis," in *Proc. IEEE Comput. Soc. PRIP Conf.*, Dallas, TX, Aug. 3–5, 1981.

[16] J. M. Lester *et al.*, "Local transforms for biomedical image analysis," *Comput. Graphics Image Proc.*, vol. 13, pp. 17–30, 1980.

[17] M. Kunt, "Source coding of X-ray pictures," *IEEE Trans. Biomed. Eng.*, vol. BME-25, Mar. 1978.

Thierry Pun (S'79) was born in Paris, France, on May 14, 1956. He received the Diplôme d'Ingénieur degree in electrical engineering from the Swiss Federal Institute of Technology, Lausanne, Switzerland, in 1979.

He is with the Signal Processing Laboratory of the Swiss Federal Institute, where he is currently completing his Ph.D. work on image processing with the support of the Swiss National Fund for Scientific Research. He also serves as a Graduate Instructor in teaching. His research interest is in digital image processing applied to scene simplification for visual prosthesis. He has published several papers related to this topic.

Mr. Pun is a student member of Eurasip.