Learning to Recognise Human Faces

October 19, 1992

Dr. L.Spacek
Department of Computer Science, University of Essex, Wivenhoe Park
Colchester CO4 3SQ, Great Britain
tel. +44 206 872343, e-mail spacl@essex.ac.uk

Dr. M.Kubat
Institute of Biomedical Engineering, Graz University of Technology,
Brockmanngasse 41, 8010 Graz, Austria
tel. +43 316 82169431, e-mail mirek@fbmtds04.tu-graz.ac.at

Subject Area: Vision - Object Recognition

Abstract

Recognition of human faces is an ambitious problem, being currently attacked by psychologists, cognitive scientists, and—to a limited extent—also by AI-community. Nevertheless, computer programs solving this task are still rare. Most of them rely on the artificial neural nets, which are not used in our approach to the problem.

The presented paper reports a successful attempt to extract a reliable set of stable intrinsic features from the images by using edge-detection, boundary grouping, and boundary characterisation. Particular attention is paid to the local properties of the boundaries at junction points. No attempt to attach high-level meaning to the individual features is made.

The resulting symbolic descriptions are processed by a simple Machine-Learning program constructing a recognition scheme in the form of a decision tree. In spite of some constraints—frontal head-on view, limited training set—the results, as measured by predictive accuracy, are promising for dealing with larger numbers of individuals.

This paper is currently not under review for a journal or another conference, nor will it be submitted during IJCAI's review period.

1 Introduction

Generally speaking, the recognition and assessment of complex real-world objects through the analysis of visual information, is one of the crucial objectives of AI. The discipline of Computer Vision has developed many sophisticated techniques for extracting symbolic descriptions of images. Computer programs have been implemented that recognise simple objects under special conditions. However, traditional approaches to object recognition suffer from the great variety of possible appearances that each object may exhibit due to changes in noise, background, illumination, viewing geometry, object orientation, occlusion, and other extraneous factors (the problem of *instability*). Further shortcoming that cannot be neglected is the programming effort that must be spent to make a machine recognise even simple objects (the problem of *costs*).

Recent developments in Machine Learning—especially in the data-analysis applications of conceptual inductive learning—indicate that the process might be prone to automation. The envisaged methodology consists of the following steps:

1. Take series of images of positive and negative examples of the object in question. Process them by computer-vision techniques to obtain their characteristics, expressed by means of primitive descriptors of regions and boundaries (The Primal Sketch), paying particular attention to the perceptual stability of the selected descriptors;

2. Submit the characteristics thus obtained to a machine-learning program that will discover regularities and develop descriptions of the objects in the images;

3. Use the machine-learning output for recognition purposes.

(For a good survey of the work done so far in combining Computer Vision and Machine Learning, see [8].)

Note that this suggested methodology bypasses the well known problems of 3D object description and matching. We wish to examine the hypothesis that good quality of 2D descriptions, coupled with a learning process, can lead to reliable object recognition. We do not expect this hypothesis to hold for all domains. For the time being we decided to venture into the ambitious task of human face recognition.

The task is formulated as follows: The computer is presented with a set of images of different persons. The machine should learn (1) to discern each individual and (2) to state whether a new image, unseen in the learning phase, represents person P or not.

Though the task of face recognition is rather novel for the AI-community, some work in this field has already been done. The previous papers fall roughly into two categories:

Psychology-oriented. These have been predominant so far. They are often published in the psychology journals and are directed towards the question of what exactly it is that enables humans to discriminate between faces or to recognise faces. The relative importance of various facial features and their cognitive representations have been studied. For an example of seminal work in this field, see [12]. For the latest paper, see [2]. The references therein provide a good coverage of the field.

Computer-oriented. This includes a few practical attempts at computer recognition. Unfortunately, the results are often difficult to interpret and to compare, as they depend to a large extent on the various assumptions made by the authors, size of the set from which the recognition is being made, etc. The difficulty of the task is such that some limiting assumptions, eg. a particular viewing direction and orientation, are accepted as normal practice in this domain. This can be justified by reference to passport photographs and the fact that human subjects, too, find it surprisingly hard to recognise photographs that are upside down. The interest in the computer face recognition task is growing [3]. Jia and Nixon have noted in their recent paper [5] that neural networks appear unsuitable for identification from large face population. In their work, they have extended the feature set to be used in automatic face recognition and have developed algorithms to extract the features automatically from a single frontal view image.

The paper presented here reports on a successful attempt to apply traditional symbolic machine-learning techniques to a carefully selected set of features describing images of human faces. After a brief exposition of the feature-extraction method and the machine-learning procedure, we present experimental results and a discussion about some aspects of the research.

2 Description Stage

The key to success are such descriptions that can be easily processed by machine-learning methods and are, above all, stable.

We used as input 24 bits/pixel Red-Green-Blue monocular images with spatial resolution of 768 x 576 pixels. The images of human faces were taken from roughly the head-on view, without particular attention being paid to the lighting, distance from camera, or background. The images were then described in terms of edge-elements grouped into boundary fragments. The RGB edge-detector and the boundary grouping program will be described elsewhere. For an earlier edge-detector of this kind, see L.Spacek [13]. The final part of the description process consists of identifying informative points on the boundaries, specifically the end points and the points of significant curvature. The entire process is similar to the construction of the Primal

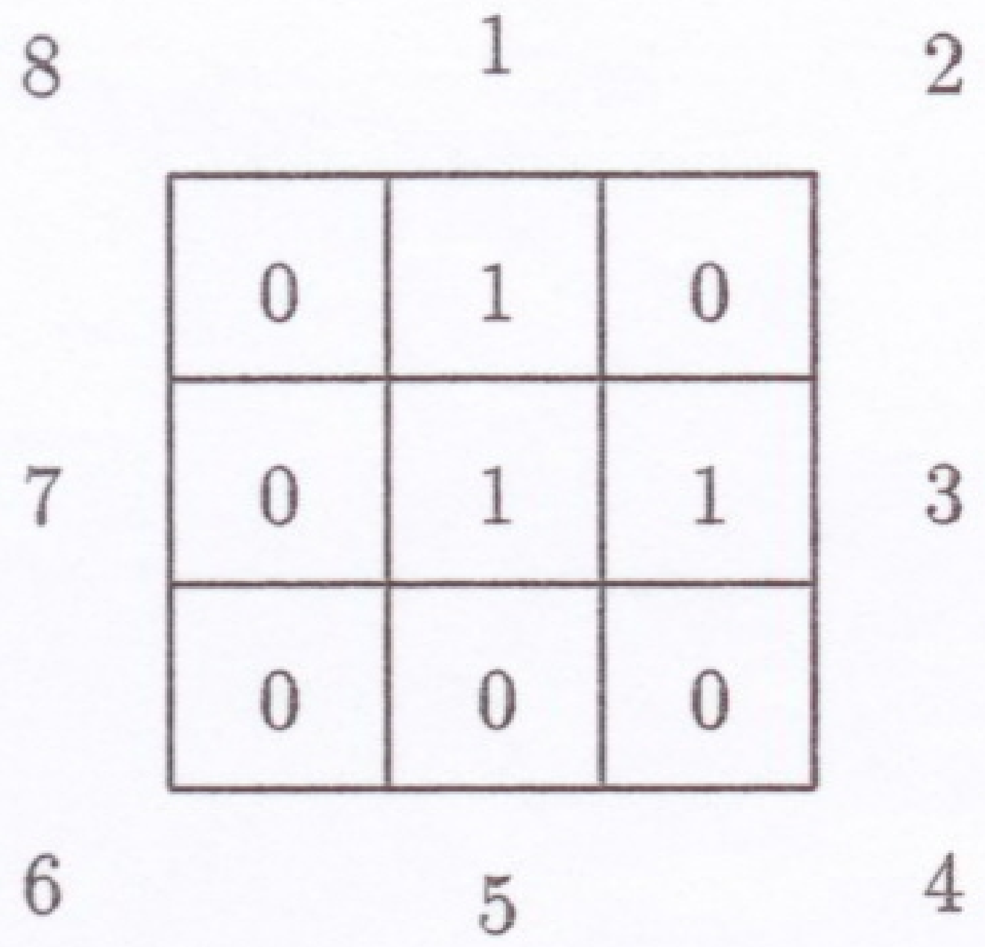

Figure 1. Occupied fields and their neighbours

Sketch, as proposed by D. Marr [7]. Other related work is that of Guo Lei [6], and, more recently, A. Noble [10] and M. Fleck [4].

The *description language* to describe the individual objects is based on the frequency of various kinds of edge curvatures. The initial image is translated into 36 numeric attributes, each giving the occurrence frequency of a particular type of a characteristic local boundary shape.

The *feature extraction* module represents the original contribution of this paper and thus deserves a more detailed description. The input of this module has the form of a matrix of fields (pixels), each containing the value '0' or '1'. '1' indicates that the field lies on a boundary—the field is 'occupied.' All other fields, labeled with '0,' are 'empty.' Note that each pixel inside the image has eight neighbours.

For each occupied field, the program finds the number of occupied neighbours. For instance in Figure 1, the middle field has 2 occupied neighbours, indicating that it lies on a boundary, 3 neighbours would indicate a 'T' crossing, 4 neighbours an intersection of 2 edges, etc. A single neighbour signals a boundary-end. For the subsequent processing, only fields with less then four neighbours were used; more complicated junctions are typically filtered out at the previous stage of boundary grouping.

In Figure 1, each of the neighbouring directions is assigned a digit, starting with '1' for north and proceeding in a clockwise direction through '8' for north-west. The combination of the digits assigned to the neighbours says whether the edge is straight or not, gives its orientation, and if it is curved, what is the orientation and size of the angle. This is the discretised representation of the 'interesting' boundary shapes which enables us to compute their frequency distributions. In Figure 1, the line is curved, the angle of the size 90 degrees pointing to SW, which is described as '1-3.' The triangular matrix in Figure 2 contains 36 blank spaces for 28 combinations of pairs of occupied neighbours and 8 directions of edge-endings (only one occupied neighbour). Junctions of three boundaries are described simply as three 'L' junctions. The same additive description could be used for more complicated junctions. This scheme has the advantage that the accidental junctions retain a measure of stability in parts of their descriptions. The program fills the blanks with data indicating the frequency of individual events. The number 0.012 in the first column of the second row says that 1.2% of the curves point to the south with the angle 90 degrees.

Each image is thus described by the 36 attributes contained in the above matrix and is assigned the classification value (typically the name of the person).

3 Learning Method

To demonstrate clearly the utility of the idea of coupling Machine Learning to Computer Vision, it is necessary to apply some well-known and simple learning algorithm. A good choice seems to be the method based on induction of decision trees, i.e. *TDIDT*—top down induction of decision trees, known also under the name of its first successful application, *ID3* [11]. The main reason is that this algorithm searches for the attributes with the rich-

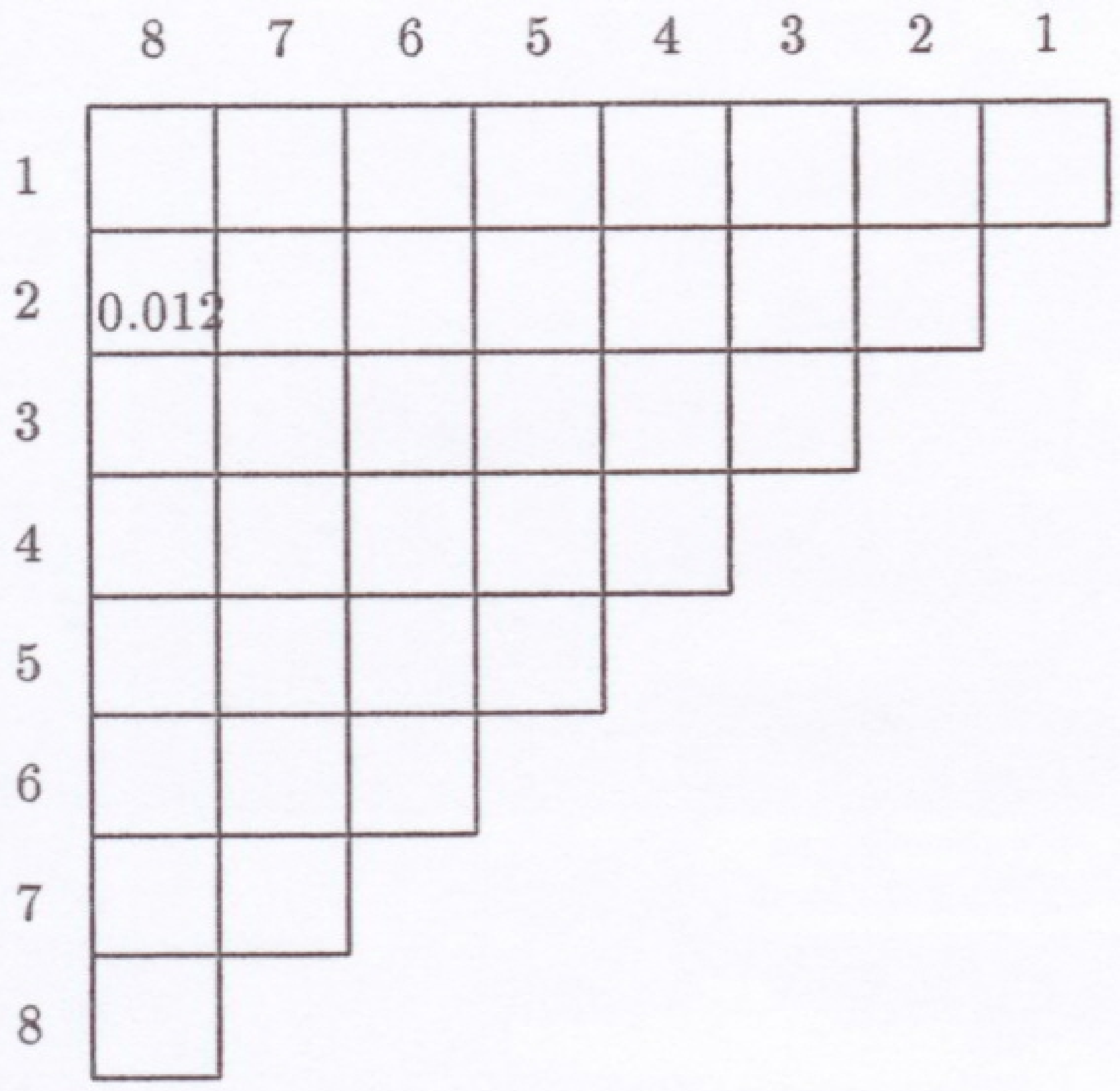

Figure 2. 36 attributes describing an image
(the diagonal contains edge-endings)

est information content—lowest entropy. The informational ordering of the attributes facilitates better understanding of the problem: mind that we wanted to assess the utility of the selected features.

From the many variants of *ID3*, we picked the simplest version, which processes symbolic data and applies an elementary method for pruning. Although also numeric versions exist, we decided not to use them. First, they are not so wide-spread and well-known outside the machine-learning community, and, second, they tend to be rather slow and we intended to carry out a lot of experiments.

We applied a simple learning mechanism, consisting of two steps: (1) data transformation from numeric to symbolic values (N/S) and (2) induction of a decision tree.

N/S-transformation.

The essence of the algorithm for the N/S-transformation consists in splitting the range of values into intervals. Each interval then represents a symbolic until there is no unused attribute left (in the last case the final subset will be assigned more than one value).

When the tree is constructed, the next step is its pruning (see [9]) to avoid overspecialized face descriptions and to discard noise. Pruning consists in cutting off those branches that are not satisfactorily grounded.

Recognition.

During the recognition phase, the system proceeds from the root to the leaves, choosing its path according to the attribute values in the respective nodes. If the person to be recognized has an attribute value that does not appear in the tree, then the closest leaf is selected. If more that one leaf is a potential candidate, then the one is preferred that has more frequently been encountered in the learning phase (has a higher a priori probability).

4 Experimental Results

The original data set, used to test the method, contained 24 images of human faces. 7 of them belonged to person 'P,' the remaining 17 images belonged to other 14 people of both sexes. It is necessary to remark, here, that 'P' complicated our task by varying his outlook—some pictures of 'P' are with glasses, some without glasses, with smile or without smile, and the like.

In order to obtain a larger set of experimental data, we have multiplied the original data set by means of 'artificial noise': for each attribute of each image, a random number $x \in< -n; n >$ (e.g. $n = 5, 10, 15$, and 20) is generated. Then, $x\%$ of the original value is added to the attribute. We have repeated the process 5 times with the same value thus obtaining 120 objects.

The usual procedure to test a new method in Machine Learning is to split the original set of examples into two disjoint subsets: one is called the *training* set, the second is called *testing* set. In our experiments, both sets were of equal size—50% of all examples.

As stated in the Introduction, two kinds of experiments were relevant for our task. First, we wanted to know to what extent the system is able to discriminate person 'P' from any other person (P against $nonP$). Second, to what extent is the system able to correctly classify any example unseen during the learning phase.

Tables 1 and 2 contain the results for both of these experiments. The parameter q is a generalization constant used during transformation N/S. Since the results were quite insensitive to the generalization constant c used by the construction of the decision tree (because a good deal of the generalization took place during the transformation), we only give results for the case where $c = 0$. Each value in the tables has been achieved as an average of 10 runs over a random split into the training and testing sets.

Table 1. Discriminating person 'P' against 'non P'

	5% noise	10% noise	15% noise	20% noise	25% noise
$q = 5$	98.2	96.8	81.0	87.7	86.0
$q = 10$	95.5	94.2	93.1	83.7	87.3
$q = 15$	99.2	97.7	96.0	86.2	91.0

Table 2. Recognition of each individual person

	5% noise	10% noise	15% noise	20% noise	25% noise
$q = 5$	88.0	79.5	72.3	55.3	48.8
$q = 10$	90.0	83.8	68.9	60.7	56.5
$q = 15$	90.5	86.7	85.2	59.7	55.8

5 Discussion

The contribution of this paper is threefold. First, the idea of applying machine-learning techniques to teach the computer to recognise complex real-world visual objects was introduced and its feasibility was demonstrated on a non-trivial real-world problem.

Second, a useful set of features to describe complex images was defined and successfully applied. Among the advantages of these variables, we would like to stress their stability. They are also relatively easy to obtain from the raw data. The nature of the selected features enable arbitrary extension of the basic set. A computer program extracting the basic features from raw visual data has been implemented.

Finally, the old challenge of automated recognition of human faces was attacked from the symbolic point of view. It is our hope that the results outlined in this paper will help to bring more attention to this exciting domain.

Human faces are often described by high-level features, such as the size and shape of the nose, type of moustache, lips, and so on. As viewed from the contemporary Computer-Vision perspective, such descriptions are very abstract and difficult to discover in the raw image. Another extreme is the use of the raw image data.

The solution presented in this paper consists in defining new features that lie—as far as the degree of abstraction is concerned—somewhere between the two extremes. The idea of the recognition of faces by mere frequencies of various points of curvature might seem bizarre. However, what matters here is simplicity and effectiveness. The former is indisputable, the latter is yet to be confirmed using a large set of images.

The first step of the research having been successfully completed, the next task is to outline the threads of the future study. We are currently using much larger samples of data and we are relaxing the problem by allowing for different sizes of the image and for elements of instability—in particular, arbitrary rotation, scaling and translation.

Obviously, new features will have to be defined. In the search for them, we again suggest the use of Machine Learning—this time to discover new concepts in the visual data. A pioneering work in this respect has been done by J.W.Bala, R.S.Michalski, and J.Wnek [1].

Since the reported results have been obtained using a very simple Machine-Learning technique, we expect, quite naturally, further improvement from more sophisticated methods that we are currently working on.

References

[1] Bala,J.W.–Michalski,R.S.–Wnek,J.: The Principal Axes Method for Constructive Induction. *Proceeding of the 9th International Workshop on Machine Learning*, Aberdeen 1992

[2] Brunas-Wagstaff,J.B. et al: Repetition Priming Follows Spontaneous but not Prompted Recognition of Familiar Faces. *The Quarterly Journal of Experimental Psychology 44A(3) (1992)*, 423–454

[3] IEE Electronics Division Colloquium: Machine Storage and Recognition of Faces. *Digest No. 1992/017* London,GB, 24th January 1992.

[4] Fleck,M.M.: Multiple Widths Yield Reliable Finite Differences. *IEEE PA-MI, vol.14,no.4 (April 1992)*, 412–429

[5] Jia,X.–Nixon,M.S.: Extending the Feature Set for Automatic Face Recognition. *Proceedings of the International Conference on Image Processing and its Applications, Maastricht, April 1992*, 155–158

[6] Lei,Guo: Level Crossing Curvature and the Laplacian. *Image and Vision Computing Vol.6,3 (1988)*, 185–188

[7] Marr,D.: Visual Information Processing: The Structure and Creation of Visual Representations. *Phil. Trans. R. Soc. Lond. B 290 (1980)*, 199–218

[8] Moscatelli,S.–Kodratoff,I.: Advanced Machine Learning Techniques for Computer Vision. In: Mařik.V.–Štěpánková,O.–Trappl,R. (eds.): *Advanced Topics in AI.* Springer Verlag, Berlin 1992, 161–197

[9] Niblett,T. Constructing Decision Trees in Noisy Domains. In: Bratko,I.–Lavrač,N. (eds.) *Progress in Machine Learning.* Sigma Press, Wilmslow (1987)

[10] Noble,J.A.: Finding Half Boundaries and Junctions in Images. *Image and Vision Computing Vol.10,4 (1992)*, 219–232

[11] Quinlan,J.R.: Induction of Decision Trees. *Machine Learning 1 (1986)*, 81–106

[12] Rhodes,G.-Brennan,S.: Identification and Ratings of Caricatures: Implications for Mental Representation of Faces. *Cognitive Psychology 19 (1987)*,473-497

[13] Spacek,L.: Edge Detection and Motion Detection. *Image and Vision Computing Vol.4,1 (1986)*, 43–56