

Learning for Feature Selection and Shape Detection

Rita Cucchiara, Massimo Piccardi, Michele Bariani, Paola Mello

Dipartimento di Ingegneria University of Ferrara,
via Saragat 1 - I-44100 Ferrara, Italy
Tel +39-532-293800 Fax +39-532-768602
e-mail: rcucchiara@ing.unife.it

Abstract. The paper proposes a general framework for shape detection based on supervised symbolic learning. Differently from other visual systems exploiting machine learning, the proposed architecture does not follow the object segmentation - feature extraction and (learning based) classification approach. Instead, an initial data-driven processing selects points of interest in the scene by means of complex features which hypothesize the presence of the target shape; hypotheses are validated by a classifier defined by a machine learning algorithm. Learning is exploited not only for defining the model, i.e. the description of the target for the classifier, but also for defining the description language, i.e. the feature set useful in generating reliable object hypotheses. The proposed architecture of visual system has been implemented for an industrial application of unstructured shape detection: examples and results are reported in the paper

1. Introduction

Computer vision community is engaging in generalizing vision paradigms and devising complete and flexible approaches to visual tasks. Many proposals of visual systems include now a machine learning stage. Machine learning is adopted under different paradigms mainly to face uncertainty associated with noise, intending noise in a broad sense (including sensor noise, cluttering, target occlusions, signal-to-symbol distortion induced by view-centered observations) [1]. However, learning is often essential for handling the model definition and providing adequate classification parameters in the case of detection of unstructured objects or shapes. In fact, the model of object targets cannot always be a priori defined in terms of geometric, topologic or other metric features (as for instance it easily performed for many handmade objects), but in many contexts only a qualitative description of the target is available. Many application fields are covered by this last framework and may take advantage of learning for defining target models: examples are recognition of hand gesture, landscape inspection, medical images analysis, and appearance-based recognition[1,2,3,4]. Another interesting example is quality inspection, where humans are very skilful at processing visual stimuli and performing classification but they do not exhibit the same ability in inferring and formalizing the rules for classification [5]. According to these considerations, we propose an architecture for unstructured and qualitatively described shape recognition that makes a substantial use of learning with the goal of defining a reliable shape model used in the recognition process. Learning affects also the image analysis process since its results are applied for setting the

description language used in the final recognition. Learning is used in conjunction with a hierarchical image analysis system based on a data-driven paradigm for performing layered parametric transformations. This proposal recalls other similar approaches, and in particular the one of Bolle and al. in [6], which describes a complete approach for object recognition based on parametric transformations on the feature spaces with the aim at providing a scene description in terms of complex or “generalized” features. In [6] parametric transformations are associated with a recognition network under a constraint satisfaction scheme for representing the object models. Instead in our proposal selection of visual features and related parametric transformations and classification are guided by the results of a symbolic learning process.

In the paper we describe the architecture of the visual system, by focusing on the layered and hierarchical image analysis block and the tree-based classifier. We present the object representation based on a hierarchical feature set together with an example of industrial application in the context of quality inspection.

2. The visual system architecture

According to the wide related literature, we aim at detecting objects in images on the basis of a set of visual features whose selection is goal-directed. Fig. 1 shows a scheme of the proposed architecture that is based on some fundamental issues: 1) the image analysis system for feature extraction providing object representation is hierarchical and layered; 2) the object representation is not performed by segmentation followed by feature measurements; on the contrary, significant features extracted from data hypothesize the presence of objects and then the feature tuples represent a possible object; 3) tuples are matched with the object model by means of a classifier which symbolically encodes the knowledge of the model; 4) the learning subsystem is based on a supervised symbolic learning algorithm which starting from a training set infers rules and the related classifier; moreover it selects the minimal and reliable set of features used in the hypothesis representation.

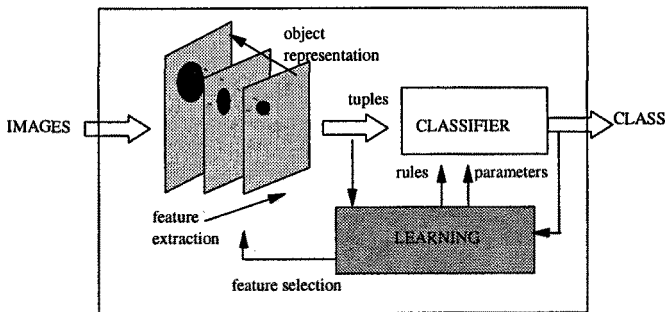


Fig. 1. The visual system architecture

The architecture defines a goal-directed overall process that is in essence data-driven at the lower levels while is strictly model-based at the highest level. The specific goal guides an initial selection of possible visual features. The result is a description of the scene in terms of features. Features are structured in hierarchical layers, with growing

levels of abstraction in order to deal with the increasing complexity of scene aspects extracted by the visual process. This approach follows the proposal of “generalized features” of [6], so that the feature computation of a higher level is carried out by means of parametric transformations as a function of other features of the same or lower levels. Therefore, as indicated in Fig. 2, we define many layered multidimensional spaces: each S_j space contains points represented by a feature tuple $V^j = \langle v^j_1, \dots, v^j_{k_j} \rangle$. Therefore features of each level accumulate evidence of salient particulars of the scene, in order to generate goal-dependent hypotheses: in the case of shape detection, hypotheses are related to the possible presence of the target shape.

A peculiar aspect of our work is that not all the possible combinations of features of different levels will be used in the classification. Instead we select only those points of the highest level identified by “high” values of semantically significant features. The object hypothesis is generated with a top-down backward process, starting from top-level features and using only those features of the lower levels generating the ones of higher levels. In practice, the object hypothesis is computed by exploiting either the previously executed visual tasks (of the bottom-up approach), or suitable anti-transformations or, eventually, also by operating further processing in that zone of the image where the attention has been focused. Finally each computed multilevel tuple of features is a possible hypothesis on the presence of target, represented as

$$O = \cup_{i=0 \dots j} V^i.$$

Many works use different level features in the classification: but if all potential tuples are considered, a possible combinatorial explosion of hypotheses calls for the use of “ad-hoc” classification algorithms or special-purpose neural network based approaches [1]. Instead, the definition of a hierarchical and structured limited set of features allowed for adopting also a general-purpose symbolic approach.

Different representations of knowledge models have been used in visual systems, such as constraint networks [6], interpretation tree search [4], rules [3], or combination of them. Our visual architecture includes a classifier which validates the object hypotheses on the bases of a decision tree inferred after a learning phase.

The multi-level representation is often redundant and generally could cause a severe and undesirable cost of the overall visual process. This is the second reason for using machine learning: the exploitation of a general purpose symbolic learning tool enables for tuning some learning parameters, in order to specialize or generalize the classifier depending on application constraints or acceptable error-rates. The produced decision tree allows for selecting a very compact reduced set of features of various levels which is enough expressive of the problem and can be used for the final hypothesis validation.

3. Hierarchical feature spaces for an inspection application

The previously proposed architecture has been adopted for an application of defect inspection on industrial metallic products. In this problem, defects appear to human eyes as elongated, roughly straight and thin shapes, very bright with respect to the neighbor background. No quantitative information is however available for distinguishing them from other similar shapes (due to sensor noise, surface roughness or surface blobs). Inspection must be carried out under UV light, since the target

shapes are visible only after a non destructive test, called MPI (Magnetic Particle Inspection) [7]. It allows for increasing the evidence of surface (and sub-surface) cracks, which can be recognized as thin long and roughly straight fluorescent strips.

The given qualitative description of the target suggests the adoption of some well assessed image analysis techniques, together with other special-purpose algorithms.

Gray level filtered images are used as input of the visual system. Features extracted by means of image processing are considered at the lowest level, so that the image space is the S0 space of the devised architecture. S1 is the space of simple primitives, such as curves or regions: in this application is the Hough space for lines obtained with the gradient-weighted Hough transform, which uses both information on the gradient magnitude and their orientation [8]. In this space some features can be extracted, such as the local peaks in the two semi-spaces H1 and H2 of the Hough space, between $[0, \pi]$ and $[\pi, 2\pi]$ respectively which manifest the presence of straight edges with opposite gradient orientation.

The S2 space collects more structured primitives: in our system is the Correlated Hough Transform (CHT) space that is defined by a suitable correlation between meaningful points of the Hough space [5,9]. The CHT is a correlation of the H1 points with adequately filtered points of H2. It has been proven to be very robust to noise and non ideality for revealing the presence of thin, straight and elongated shapes, as the considered target shape is [9]. Therefore the S2 points, because of CH, accumulate the evidence of possible targets: the visual system selects points of interests in S2, representing detection hypotheses. An hypothesis is described by a tuple of features, comprehending CH, some features extracted previously of computing CH, and others obtained by an anti-transformation in the points of interest of the image. In the application, the computed tuple is $O = (CH^2, H1^1, H2^1, H2a^1, Tk^1, NP^0, Va^0, Ga^0)$, where the ceiling number indicates the level of the feature space. These features are briefly described as follows:

- CH is a “high” local peak in S2; it is the most salient feature computed according to the target model [9];
- H1 is the gradient weighted Hough value in $H1 \subset S1$ at the same coordinates of the point CH in S2; it indicates the “rectilinearity” of the first target edge, by supposing the target brighter than the background[9];
- H2 is the highest local peak in $H2 \subset S1$, computed in a neighbour of the point where the second peak (representing the second straight edge) should be found, in the case of an ideally straight defect;
- $H2a \in S1$ represents the average value of the previous considered neighbour and accounts for possible not-ideality both in straight shape and in thickness.
- $TK \in S1$ is the mutual Euclidean distance between the ideal edges of H1 and H2. It is the upper bound of the shape thickness. It is used also for delimiting a focus of attention in the image space;
- $NP \in S0$ is the number of voting points of the image space which are transformed in the point of the Hough space corresponding to H1. It estimates the edge length.
- $VA \in S0$ is the average vote of the voting points, i.e. the average Gradient of each point voting for the point corresponding to H1;
- $GA \in S0$ is the average gradient of the image: differently from the others, it is a global feature of the whole image. It should be used as a corrective weight: images

containing cracks should have a “high” average gradient but not “too high”, since for the MPI process the fluorescence is collected only in the zone of a possible crack and is not distributed on the surface.

It should be noted that the tuple representing the object has been computed starting by the most salient features, CH. This limits the number of possible hypotheses extracted from images, starting only from features of lower levels: for instance, the single presence of a high value of H1 is not enough for suggesting the presence of the target, since high values of H1 (without being supported from high values of other features) can be measured also in presence of spurious fluorescent blobs or other shapes not thin enough to be classified as a defect.

As a practical example, we report some measures on inspected images in Fig.2. They represent some particulars of workpieces with or without defect. Table 1 shows their correspondent feature values, computed on images with a gray level represented with a real number in the range [0,1].

Table 1. Tuples for images of Fig. 2

	CH	H1	H2	H2A	NP	VA	TK	GA $\times 10^3$
I38	12.33	0.97	3.02	12.64	21	0.088	5	46.2
I34	13.63	1.00	3.02	12.51	21	0.084	5	44.9
I12	393.29	7.04	15.76	55.82	17	1.006	2	60.66
I17	8.44	0.59	2.72	14.10	17	0.085	3	45.44

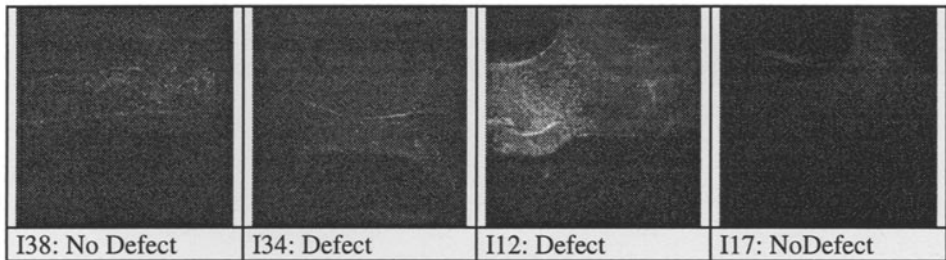


Fig. 2. Example of images

4. The automatic learning subsystem

The highest level of the visual process addresses the decision making task with reference to the knowledge base on the target. Its implementation depends on the representation used for the target model: it can be encoded within the classifier, as in the case of connectionist systems, or can be symbolically defined by means of rules, decision trees, semantic nets and other symbolic techniques exploiting predicates on features.

Our visual system includes the Quinlan's C4.5 classifier based on learning by example algorithms which automatically generate both the decision tree and, after an adequate pruning operation, also a knowledge representation as a set of production rules [10]. It has been selected after a comparison with other approaches based on different paradigms, since it resulted very suitable for managing both symbolic and continuous values in accordance with the application requirements [10,11].

Since the goal is to validate hypotheses on the presence of the target shape (i.e. the defect) by means of the feature tuples, we address a classifier based on two classes of interest, namely *Defect* and *NoDefect*.

```

C4.5 [release 8] decision tree generator - Read 317 cases (8 attributes)
Decision Tree:
CH > 78.6283 : Defect (29.0)
CH <= 78.6283 :
  CH <= 8.86457 : NoDefect (133.0)
  CH > 8.86457 :
    GA > 47.0579 : NoDefect (98.0)
    GA <= 47.0579 :
      CH > 17.7267 : Defect (18.0)
      CH <= 17.7267 :
        GA > 45.6994 : NoDefect (13.0)
        GA <= 45.6994 :
          GA <= 44.6796 : NoDefect (3.0)
          GA > 44.6796 :
            H2a > 12.1436 : Defect (15.0)
            H2a <= 12.1436 :
              H2a <= 11.4278 : Defect (5.0)
              H2a > 11.4278 : NoDefect (3.0)

```

Fig. 3. The decision tree generated by C4.5 on the training set

For each image, we considered the points of the S2 space having CH value higher than a given percentage (75%) of the maximum CH value, as possible positive hypotheses of the presence of the target shape. The used training set is composed of 317 tuples (67 Defect (21.1%) - 250 NoDefect (78.9%)). In order to train the classifier in the ambiguous region, the training set was composed so as to contain also NoDefect objects exhibiting high CH values, that are hard to be classified correctly .

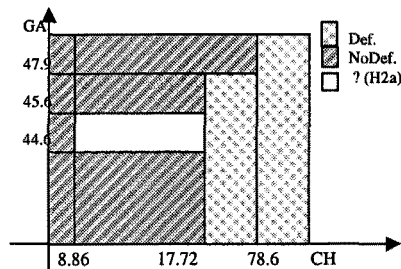


Fig. 4. Classification based on CH and GA features

Fig. 3 shows the decision tree produced by C4.5. The test set was generated by Cross-Validation, extracting one object at a time from the training set and submitting it to the classifier obtained from the remaining examples. Table 2 reports error results both in the training set and in the test set. In this last case also a partition between *false-identification* errors in the classes of Defect and NoDefect is provided. False-identification occurs when the decision tree assigns an element to a class i when it belongs to a class $j \neq i$ [12]. In the case of quality inspection the false identification in the NoDefect class is more critical, since it indicates the percentage of target which actually are not detected. Instead, the false identification rate in Defect class increases only the percentage of rejected workpieces. However it must be limited because it affects the final production efficiency.

Table 2. Results on the training and test sets

	Training Set	Test Set		
	Overall error(%)	Overall error(%)	Defect error(%)	NoDefect error(%)
C4.5 tree	0.0	2.8	4.5	2.3
C4.5 tree (no GA)	6.0	12.9	44.8	3.8

As a first important result, the use of a general-purpose machine learning tool provides a compact and efficient decision tree, by suggesting the use of specific numeric thresholds on the features and the related unary predicates. The decision tree of Fig. 5 is actually adopted by the classifier used for validating target hypotheses.

The results of learning validate the use of a hierarchical image analysis sub-system: obviously most of the predicates are carried on the highest level feature, CH, which is the most significant one; nevertheless the points of interest are represented by multi-level features. While the importance of the highest level feature, CH, (actually computed for the specific application) can be intuitively accepted, an important and less expected role is played by the average gradient of the image (GA). This feature takes into account both the image noise and the evidence of the defect with respect to its background; it turns out to be more valuable than a simple corrective weight factor since the relation between CH and GA variations and the Defect-NoDefect classification is not linear. This is described in Fig.6 by a graphical representation of the decision tree in the simplified feature space of CH and GA only.

The inner white rectangle indicates the ranges of CA and GA where the correct classification needs the information on other features (such as H2a).

Without the low-level GA feature the learning and the consequent classification obtain considerably worst results. In this case the decision tree turns out to be well more complex and specialized than the previous one. It takes into account six features of seven available, instead of only three of the tree of Fig.4. Moreover it exhibits a non negligible error-rate, as shown in Table 2.

Furthermore, a valuable contribution of this work is in the use of learning together with a layered image analysis subsystem: learning on tuples constructed starting from features of the highest level space handles for setting the minimal feature set useful for hypothesis description. In particular the decision tree suggests the use only of three features, one for each space level. The other five values in the tuple are not used by the classification tree. This is an important feedback on the vision process which enables for a priori avoiding measuring all those features, thus reducing computational load.

In this way the visual system used in a real on-line working environment is able for performing shape detection in a limited amount of time and the whole quality inspection process is provided in real-time. Several images of the workpieces are acquired so as to inspect different object poses, and a piece is classified as defective if at least an hypothesis on the defect presence is validated by the classifier. Moreover, the learning algorithm is available on-line for automatically updating the classifier in case of whichever changing of the working environment .

5. Conclusions

This work presented a visual architecture for detecting target shapes in images, based on a machine learning tool used with a hierarchical and layered image analysis system. The hierarchical structure of the image analysis block allows for maintaining the explicit relationships between features describing objects and it is exploited in two directions: a first bottom-up direction is used for focusing the attention; a second step in the opposite direction from the highest level down to the pixel level is used to compose a tuple representing one hypothesis. The hypothesis evaluation is carried out with decision trees automatically inferred from a training set.

The use of automatic learning is exploited for selecting the smallest subset of meaningful features and for easy updating the classifier when the external condition change. Learning showed that some of the objects features, even if they appeared to be salient, were not significant for the specific classification task, and therefore can be ignored. This feedback from the classification systems to the low-level image processing allowed for reducing the computational load and improving the reliability of the vision task that has been applied in a real industrial inspection process.

References

1. Cho, K., Dunn, S. M.: Learning shape classes. *IEEE Trans. on PAMI* **16** (1994) n. 9 882-887
2. Draper, B. A., Brodley, C. E., Utgoff, P. E.: Goal directed classification using Linear Machine Decision tree. *IEEE Trans. on PAMI* **16** (1994) n. 9 888-893
3. Pellegretti, P., Roli, F., Serpico, S., Vernazza G.: Supervised learning of descriptions for image recognition purposes. *IEEE Trans. on PAMI* **16** (1994) n. 1 92-98
4. Murase, H., Nayar, S. K.: Learning by a generation approach to appearance based object recognition. *Proc. of 13th ICPR, Vienna* **1** (1996) 24-30
5. Cucchiara, R., Filicori, F., Andreetta, R.: Detecting micro-cracks in ferromagnetic material with automatic visual inspection. *Proc. of QCAV95 Int. Conf. on Quality Control by Artificial Vision*, Le Creusot France, (1995) 19-24
6. Bolle, R. M., Califano, A., Kjeldsen, R.: A complete and extendable approach to visual recognition. *IEEE Trans. on PAMI* **14** (1992) n. 5 534-548
7. Newman, T. S., Jain, A.K.: A Survey of automated visual inspection. *Comp. Vision and Image understanding* **61** (1995) n. 2 231-262
8. Illingworth, J., Kittler, J.: A Survey of the Hough transform. *Comput. Vision Graphics, Image Process.* **43** (1988) 221-238
9. Cucchiara, R., Piccardi, M.: Eliciting visual primitives for detecting elongated shapes, 3rd Int. Workshop on Visual Form Capri, Italy, (1997) (to appear)
10. Quinlan, J. R.: C4.5 Programs for machine learning. Morgan Kaufmann Publ. (1993)
11. Bariani, M., Cucchiara, R., Piccardi, M., Mello, P.: Data mining for automated visual inspection, *Proc. of PADD97, London, UK*, (1997) 51-64
12. Haralick, R., Shapiro, L.: *Computer and Robot Vision* vol. I, Addison Wesley (1991)