Recognition I

Applying VC-dimension Analysis To Object Recognition *

Michael Lindenbaum Shai Ben-David
Computer Science Department, Technion
Haifa 32000, ISRAEL

Abstract

We analyze the amount of information needed to carry out various model-based recognition tasks, in the context of a probabilistic data collection model. We focus on objects that may be described as semi-algebraic subsets of a Euclidean space, and on a wide class of object transformations, including perspective and affine transformations of 2D objects, and perspective projections of 3D objects. Our approach borrows from computational learning theory. We draw close relations between recognition tasks and a certain learnability framework. We then apply basic techniques of learnability theory to derive upper bounds on the number of data features that (provably) suffice for drawing reliable conclusions. The bounds are based on a quantitative analysis of the complexity of the hypotheses class that one has to choose from. Our central tool is the VC-dimension, which is a well studied parameter measuring the combinatorial complexity of families of sets. It turns out that these bounds grow linearly with the task complexity, measured via the VC-dimension of the class of objects one deals with.

1 Introduction

Object recognition is one of the fundamental tasks of Computer Vision, and has drawn a lot of research attention. In this task, one observes an image, looks for evidence to the presence of known objects in it, and tries to find their identities and positions. The performance of the recognition methods is usually estimated by implementing and testing them on simulated and real data (see [MH90] for an example to such extensive experimentation). This evaluation methodology, while very reliable, makes the comparison between the methods difficult, because, typically, they are tested only on different small sets of test objects, and use different kinds and amounts of data, collected with different precisions.

We are interested in the amount of data collected from the image, that is required to succeed in a recognition task. Intuitively, we need more data if the recognition

*This work was supported by the Technion fund for the promotion of research and by the Smoler research fund

Lecture Notes in Computer Science, Vol. 800
Jan-Olof Eklundh (Ed.)
Computer Vision – ECCV '94

procedure is required to discriminate between object instances that are visually similar, and if more alternatives are allowed by the possible instance specification. In this paper these intuitive observations are quantified. We consider a random measurement model, which places data features in the image according to a certain probability distribution, that depends on the real object present in the scene. We derive an upper bound on the number of features required to succeed in the recognition task with a certain confidence.

The bounds are derived relying on the observation that the recognition task is equivalent to a learning task, in which one tries to learn a subset of some space, by observing samples of this space. We use well known results from the *Computational Learning Theory* field, which investigate, among other issues, the number of examples required to learn. The elegant PAC learning framwork (described in the next section) states that the number of samples required to learn is not higher than a certain threshold, which grows with the accuracy of the hypothesis, the required confidence, and a certain parameter, associated with the class of allowed hypotheses, and known as the VC-dimension. The VC-dimension parameter, introduced by Vapnik and Chervonenkis is associated with a class of sets, (or more generally with a class of functions,) and is a combinatorial measure that characterizes its generality [VC71].

we use these results to find the number of data features required to succeed in a recognition task. For the simpler localization problem, where the object's identity is known and only its position is to be found, we consider the class of instances associated with one object and some class of transformations, and find the VC-dimension parameter associated with it. Interestingly, the analysis and its results are independent of the particular object considered, and the derived VC-dimension parameter depends only on the object's complexity and the class of transformations. Deriving the VC-dimension associated with a recognition task is an easy extension. A straightforward application of the bounds given in [BEHW89] gives now the number of data features which guarantees that the hypothesized instance is not far from the true instance. More concrete assertions, such as that the recognition results is "true" easily follow.

In a way, the number of measurements required for recognition quantifies the *fundamental difficulty* of the recognition task, that is, the ability to succeed in a recognition task regardless of its computational cost. This should not be confused with the *computational difficulty*, which quantifies the time (or space) needed to process these measurements.

The *fundamental difficulty* of recognition tasks was considered before in a few papers: Lindenbaum analysed the effect of the object's shape on the recognition difficulty and quantified it by two parameters, Self-similarity and similarity, that characterize the difficulty of localization and recognition, respectively [L92]. Another aspect of the *fundamental difficulty* of recognition is considered by Grimson and Huttenlocher, who analysed the possibility that a subset of "noise data features" will give a false evidence for the presence of an object in the scene [GH91]. Similar methods allow them to analyze the more general affine case [GHJ92]. More recently Maybank analyzed the cross ratio invariant function and was able to predict the probability of "a false alarm" as a result of noisy image features [M93]. These results are complementary to ours as they analyze another reason, the image clutter, for recognition failures.

The paper is divided into two major parts: explaining the relation between learning and recognition, and calculating the VC-dimension associated with certain recognition tasks. We start with a brief characterization of learnability using the VC-dimension parameter. Then, in the next section, we present a mathematical formulation of the recognition process, composed of a data collection stage and an interpretation stage and show the relation between recognition an learning. The second part is devoted to calculating the VC-dimension associated with certain recognition tasks. In this shorter version, we analyse in detail only the task of localizing 2D semi-algebraic objects under the Affine, Similarity, and Euclidean transformations. Much of the abstract mathematical background of this paper, without the interpretation we give here was considered already in [BL93] and in [GJ93]. More results, as well as the proofs to the theorems presented here, may be found in [BL93] [LB93].

2 Learnability and the VC-Dimension

Given a collection, $\mathcal{K}$, of subsets of some base set, X, and a measure of difference between the members of $\mathcal{K}$, a set of points $\{x_1, \ldots, x_n\} \subset X$ is said to ϵ*−pin down* $\mathcal{K}$, if, for every pair of sets $A, B \in \mathcal{K}$, if $A \cap \{x_1, \ldots, x_n\} = B \cap \{x_1, \ldots, x_n\}$ then the difference between these members of $\mathcal{K}$ is at most ϵ.

It is evident that the size of such 'pinning down' sets, as well as their number, depends upon the family $\mathcal{K}$ of sets. The theory of computational learnability formalizes this issue within the framework of Valiant's PAC learning model. In that model the family of sets $\mathcal{K}$ is usually called a 'concept class' and its members are 'concepts'. The model assumes the existence of some probability distribution P over X. This probability plays a double role: First, the difference between concepts is specified as the P-probability of hitting their symmetric difference. Second, the 'fraction' of pinning-down n-tuples (among all n-tuples of points of X) is measured by the probability of picking such a tuple by i.i.d. sampling n-many times according to P.

A class $\mathcal{K}$ is called *PAC-learnable* (or just 'learnable') if, for every positive ϵ, δ, there exists a finite number m (depending upon these parameters) such that for every probability distribution P over X, the P^m-probability of picking an m-tuple that ϵ-pins down $\mathcal{K}$ exceeds $(1-\delta)$. The cornerstone result of the theory of PAC-learnability is a complete analysis of the learnability of concept classes in terms of a purely combinatorial parameter – the Vapnik-Chervonenkis dimension of the class.

Definition 1: *[Vapnik-Chervonenkis Dimension] Let X be some set and $\mathcal{K}$ a collection of its subsets.*

- *We say that* $\mathcal{K}$ shatters *a set $A \subseteq X$, if, for every $B \subseteq A$, there exists some $C \in \mathcal{K}$ such that $C \cap A = B$.*
- *The* Vapnik-Chervonenkis Dimension *(in short, VC-dim) of $\mathcal{K}$ is the maximum number d such that $\mathcal{K}$ shatters a set of size d. (If $\mathcal{K}$ shatters sets of unbounded size, we say that its VC-dim is ∞).*

Example: Let X be the unit interval and $\mathcal{K}$ be the collections of all its subintervals whose length is 0.1. I.e., $\mathcal{K} = \{[a, a+0.1] : \ 0 \leq a \leq (1-0.1)\}$. It is not hard to realize that $\mathcal{K}$ shatters every pair of points in [0.1, 0.9] which are at most 0.1 apart. On the other hand, $\mathcal{K}$ shatters no subset A of the interval whose cardinality exceeds 2. It follows that VC-dim$(\mathcal{K}) = 2$.

We state now the upper bound given by Blumer et. al. [BEHW89] on the number of examples required for pinning down a set.

theorem 1 *[[BEHW89]] If $VC - dim(\mathcal{K}) = d$ then, for every positive ϵ and δ, if*

$$m \geq \max\left(\frac{4}{\epsilon}\log\frac{2}{\delta}, \frac{8d}{\epsilon}\log\frac{13}{\epsilon}\right)$$

then, for every probability distribution P over X, the P^m-probability of picking an m-tuple that ϵ-pins down K exceeds $(1-\delta)$.

3 Learning and recognition

This section discusses the relation between learning and recognition, and shows that in a proper setting, recognition tasks are equivalent to learning tasks in the sense that an object is recognized (or localized) if some related concept class is PAC learned with a certain prediction power.

We consider recognition processes that are composed of a data collection stage followed by an interpretation stage. In the first stage data features are collected in random locations, independently, and according to fixed distribution. In the second stage the data collected is combined with prior knowledge, and is interpreted, to yield an hypothesis on the identity and pose of the object in the scene. These stages are described in the next two sections.

3.1 The data collection stage

In Vision scenarios, information is usually obtained from the observed object's edges in an image, and is usually associated with some location error. Depending on the imaging conditions, one is able to extract either full portions of the boundary, just some boundary points, or boundary points associated with accompanying parameters, like, for example, the boundary slope. The simple model, suggested in the following lines, is not claimed to cover all situations in computer vision. It addresses, however, important issues such as the uncertainty on the observed part of the object and the inaccuracy of the measurements. We consider data features that are randomly drawn in the neighborhood of the object boundary. In the simple case, where only boundary points associated with inaccuracy Δ are available, we assume that they are independently sampled inside the Δ-neighborhood of the object boundary. Let ∂O_t be the boundary of the instance of the object O, after a transformation t. Then, we assume that the collected data are independently sampled, according to a uniform distribution, inside

$$O_t^{\Delta} = \{\ r \mid \exists s \in \partial O_t \ s.t. \ ||s - r|| < \Delta\}, \tag{1}$$

to which we refer as "extended boundary". More complicated data collection models, which include arbitrary but bounded sampling distributions and data features which include boundary slope measurements, are considered in the full version and elsewhere.

3.2 The interpretation stage

Let O_t denote an instance of an object O which corresponds to this object after transformation $t \in T$. Let H be the set of possible hypotheses, which, in the model based setting, may contain instances of different objects under different transformations. We refrain from referring to any particular method for inferring the hypothesis. The only assumption taken is that the interpretation stage may draw any hypothesis that is consistent with the data. That is, for M being the data set, the algorithm may draw any hypothesis in $\{h | M \subset h;\ h \in H\}$.

An error measure

We treat all recognition tasks uniformly and consider them successful if a special error measure, defined below, between the true object and the hypothesized one, is guaranteed to be lower a threshold value. For O_t being the true object that is present in the scene and $W_{t'}$ being some hypothesized instance, the error associated with this hypothesis is defined as the normalized difference between the volumes of the corresponding extended boundaries.

$$E(O_t, W_{t'}) = \frac{Vol(O_t^{\Delta} \setminus W_{t'}^{\Delta})}{Vol(O_t^{\Delta})} \tag{2}$$

This error measure agrees with the intuitive meaning of recognition and localization. High localization accuracy, for example, implies that the boundaries of the true object and the hypothesis are very close, and leads to a small difference between the corresponding extended boundaries. Low localization accuracy, on the other hand, allows larger error. Inferring the recognition performance for more traditional distance measures, like the Hausdorff metric for similarity, the maximal distance between corresponding points for localization, and the same/different binary distance for discrimination may be done, as was shown in [L92], and will be briefly described in the sequel.

The uniform recognition accuracy measure may be used to specify recognition success in the more familial forms, by setting the maximal error, for which the hypotheses is still considered successful as follows:

1. **discrimination problem:** Let V and W be two planar objects and T a class of transformations. Let
$$e_0 = \max_{t,t' \in T} E(V_t, W_t').$$
Clearly, requiring a recognition accuracy better than e_0 guarantees that no instance of W is drawn as an hypothesis if the true object is an instance of V.

2. **localization problem:** Let V be a planar object and T a class of transformations. Consider any distance measure $D(\cdot, \cdot)$ between two object instance, and denote a localization procedure successful if, for the hypotheses drawn,

the distance between the true object V_t and the hypothesis $V_{t'}$ is d_0 or smaller. (The value d_0 may be adjusted arbitrarily according to the localization precision required.) Let

$$e_0 = \max_{t,t' \in T \; ; \; D(V_t, V_{t'}) > d_0} E(V_t, V_t').$$

Requiring a recognition accuracy better than e_0 guarantees that no instance of V which is d_0-far from the true instance is drawn as an hypothesis.

For both tasks, the question we are interested in will now be

How many measurements are needed to guarantee, with a certain confidence $1 - \delta$, *that all hypotheses that are at least* e_0*-far from the true object instance are rejected ?*

3.3 Learning and recognition

Now, the equivalence between recognition tasks, and PAC learning should be apparent: For the localization task, for example, let $\{O_t | t \in T\}$ be a set of instances associated with one object O and a class of instances T. To every instance from this set, associate a concept identical to the extended boundary.

$$O_t \longleftrightarrow O_t^{\Delta} \tag{3}$$

$$\{O_t | t \in T\} \longleftrightarrow C_{T^{\Delta}}(O) = \{O_t^{\Delta} | t \in T\} \tag{4}$$

Every data feature extracted from the object boundary provides a (positive) example to the corresponding concept. Learning a concept in $C_{T^{\Delta}}(O)$ with an accuracy better than e_0 means that all concepts in the class, associated with a symmetric difference greater than e_0, are not consistent with the examples. Note however, that according to our measurement model, the density is zero everywhere except inside the concept itself. Assuming further that the distribution is uniform within the extended boundary, implies that the recognition error (2) is also smaller than e_0, and that the recognition task is successful.

In other words, if the learning algorithm provides an hypotheses that is e_0-close to the extended boundary associated with the true object, then the recognition task is successful, in the sense that the object instance associated with the extended boundary hypotheses cannot be an instance of the wrong object (in discrimination) and/or cannot be too far from the true instance (in localization and recognition).

While the PAC learnabily results usually hold for arbitrary distribution, we will assume that the data features are placed according to a uniform distributions densities. The reason for that is the need to establish a relation between the recognition accuracy measure $E(V_t, W_{t'})$ and the symmetric difference $V_t^{\Delta} \Delta W_{t'}^{\Delta}$ Induced by the sampling density. This cannot be achieved by all distributions: Consider for example a distribution that is concentrated in a single point. The learning performance in this case will be excellent as the density weighted symmetric difference and the associated prediction error will be null after one example. The knowledge about the identity or location of the object will, however, be poor because completely different hypotheses can be consistent if they share one point with the true object (either inside or outside).

4 Localization - The VC-dimension of transformed Semi-Algebraic sets

4.1 Introduction and summary of results

To know how many data features are sufficient to guarantee that every consistent hypothesis is e_0-accurate with confidence $1-\delta$, we may now use the bound in theorem (1), and therefore we now proceed to calculating the VC-dimension of the $C_{T^\Delta}(O) = \{O_t^\Delta | t \in T\}$ concept classes. This section considers two dimensional objects which are semi-algebraic sets of degree (k, m), and instances of them created by Perspective, Affine, Similarity and Euclidean transformations. (3D objects and instances of them are considered elsewhere.)

Definition 2: *A* semi-algebraic open set *of degree* (k, m) *in* $\mathbb{R}^n$ *is a set that can be represented as a boolean combination of* k *sets of the form* $\{\bar{x} \in \mathbb{R}^n : f_j(\bar{x})\, Q\, 0\}$ *where the functions* f_j *are real polynomials of maximal degree* m*, and* Q *is one of the relations* $\leq, =, <$.

The class we consider here is very rich: besides polynomial objects it also contains combinations of them which include, e.g., polygonal objects (which, for k being the number of polygon sides, are semi algebraic sets of degree $(k, 1)$). The family of Semi-Algebraic sets is parametrized, meaning that the class of objects considered is actually not limited.

To every such object instance, a concept, equal to the extended boundary, is associated. For V denoting the semi algebraic object and $t \in T$ denoting some transformation, the corresponding object instance is denoted by V_t, and the corresponding concept is denoted by $(V_t)^\Delta$. Considering one object V, and a class T of transformations, we consider the class of concepts $C_{T^\Delta}(V) = \{V_t | t \in T\}$. In this section such concept classes, denoted just as "classes of transformed objects", will be analyzed, and upper bounds on their VC-dimension will be found, thereby providing the parameter needed to determine the amount of data required to localize the object with the required precision and confidence.

Summary of the results

For planar semi-algebraic objects of degree (k, m) and for Translation, Euclidean, Similarity, Affine and Perspective transformations groups, we find upper bounds on the VC-dimension of the corresponding concept classes, which are linear in the number of transformation parameters and logarithmic in the complexity of the object. In particular, when the object complexity increases, we show the following asymptotic bounds on the VC-dimension of the classes of *extended boundaries of transformed semi algebraic sets.*

$$
\begin{aligned}
B_{\mathcal{T}^\Delta}(V) &= 62 \log(km) \\
B_{\mathcal{E}^\Delta}(V) &= 124 \log(km) \\
B_{\mathcal{S}^\Delta}(V) &= 155 \log(km) \\
B_{\mathcal{A}^\Delta}(V) &= 186 \log(km) \\
B_{\mathcal{P}^\Delta}(V) &= 712 \log(km)
\end{aligned}
$$

4.2 A sketch of the derivation for the affine case

A partition of the transformations set

Generally, much of the analysis is similar for two and three dimensional objects, and for general transformation class. Therefore, we consider general subsets of the general real space $\mathbb{R}^n$ and a general transformation class, denoted T, in much of the analysis. We refer to the particular dimensions and transformation class, when needed.

We would like to find now the VC-dimension of the class $C_{T^\Delta}(V) = \{(V_t)^\Delta : t \in T\}$ where V is some semi-algebraic object of degree (k, m) in $\mathbb{R}^n$, and T is some transformations class. First, observe that any object V induces a mapping of points of $\mathbb{R}^n$ to subsets of T. That is, every point $\bar{x} \in \mathbb{R}^n$ is mapped into the subset of transformations $K_{\bar{x}}^V = \{t \in T : \bar{x} \in (V_t)^\Delta\}$. Note that this mapping is dual to the mapping from members of T to subsets of $\mathbb{R}^n$ defined by mapping t to the set $(V_t)^\Delta$

Before we proceed, let us introduce some further notation: Given a set S, for every subset, $A \subseteq S$, let $\theta_A(x_i)$ be the indicator function:

$$\theta_A(x_i) = \begin{cases} 1 & x_i \in A \\ -1 & else \end{cases}$$

Let the exponent notation R^1 and R^{-1} denote a subset, $R \subseteq T$, and its complement $T \setminus R$, respectively. Consider now the set of points $S = \{x_1, x_2, \cdots x_N\}$ in $\mathbb{R}^n$. Fixing an object set $V \subseteq \mathbb{R}^n$ every point subset $A \subset S$ corresponds to a subset of the transformations set T

$$W(A, S) = \bigcap_{i=1}^{N} K_{x_i}^{\theta_A(x_i)}$$

(In an attept to simplify the notation, we have dropped the superscript V from the sets K_x^V). The following claim is straightforward from the definitions:

Claim 1 *For any $A \subseteq S$ and $t \in T$,*

- *$A = (V_t)^\Delta \cap S$ iff $t \in W(A, S)$.*
- *For $t \in W(S, A)$ and $x \in S$, $t \in K_x^V$ iff $x \in A$.*
- *For any object $V \subseteq \mathbb{R}^n$ and a family T of transformations, the class of transformed objects, $C_T(V)$, shatters a set $S \subseteq \mathbb{R}^n$ iff neither of the members of $\{W(A, S) : A \subseteq S\}$ is empty.*

We have therefore reduced the calculation of the VC-dimension of classes of transformed images to counting the number of non-empty $W(A, S)$ sets of transformations. We apply this reduction in our subsequent derivations.

Parametrizing the transformations

The next step we take is to represent T parametrically, with parameters forming some parameter space $\mathbb{R}^k$. We restrict our attention now to 2D objects. A linear affine transformation on $\mathbb{R}^2$ is defined by a pair, (A, b), where A is an 2×2 matrix

and $b \in \mathbb{R}^2$. Such a transformation $H = (A, b)$ acts on $\bar{x} \in \mathbb{R}^2$ by $H(x) = A\bar{x} + b$. For non-singular transformations, the inverse transformation, denoted $H' = (A', b')$, always exists, and its parameters, the components of A' and b' will be used to represent the transformation. Clearly, the parameter vector, denoted $\bar{t}$, is a point in the affine parameter space $\mathbb{R}^6$. Now, $K^V_{\bar{x}}$ will also denote the set of parameters that correspond to all transformations t for which V_t^{Δ} includes the point $\bar{x}$. Other transformation subsets, such as the $W(A, S)$ subsets will also denote parameter subsets.

The parameter sets $K^V_{\bar{x}} \subseteq \mathbb{R}^6$, which depends on the particular measurement $\bar{s}$, will be considered now. A sufficient and neccessary condition for a point $\bar{x} \in \mathbb{R}^2$ to be inside the transformed object is that there will be a point $\bar{s} \in \mathbb{R}^2$ on the boundary ∂V_t close enough to it. Applying the inverse transformation on this point $\bar{s}$, $\bar{s'} = H'(\bar{s}) = A'\bar{s} + b'$ must be in the original (nontransformed) semi-algebraic set ∂V. Therefore,

$$K^V_{\bar{x}} = \{\bar{t} = (A', b') \,|\, \exists \bar{s} \;\; s.t. \; A'\bar{s} + b' \in V \;;\; ||\bar{x} - \bar{s}|| < \Delta \;\} \tag{5}$$

We shall show now that $K^V_{\bar{x}}$ is also a semi-algebraic set (in the parameter space $\mathbb{R}^6$), albeit of higher degree. To do that we must eliminate the variable $\bar{s}$ out of the polynomials that specify this paramter set, a task that has naturally draw a lot of attention in the field of Logic.

Logic Formulae, quantifier elimination, and Collins decomposition

Is the context of Logic, one is often interested in finding the range of variables for which some formula is true. One way to formulate this task is to consider *atomic formula* of the form $\{x|f(x) > 0\}$ and $\{x|f(x) = 0\}$, and more complex formulae that are boolean combinations of them. The later may include also quantified variables and are commonly written in a *standard prenex formula*

$$(\mathbf{Q}_{k+1}x_{k+1})\,(\mathbf{Q}_{k+2}x_{k+2})\,\ldots\,(\mathbf{Q}_r x_r)\,\Phi(x_1, x_2, \ldots, x_r) \tag{6}$$

where $\Phi(x_1, x_2, \ldots, x_r)$ is a quantifier free formula, $1 \leq k \leq r$, and each $(\mathbf{Q}_i x_i)$ is either an existential quantifier $(\exists x_i)$ or a universal quantifier $(\forall x_i)$. Assume that the variables $x_1, x_2, \ldots, x_r$, divided into *free variables* $x_1, x_2, \ldots, x_k$ and *quantified variables* $x_{k+1}, x_2, \ldots, x_r$ (denoted also just by *quantifiers*), are real. A standard formula specifies a set of the k *free variable* known as a *Tarski set*, for which the formula is true. For a quantifier free formula (that is for $k = r$,) a *Tarski set* is clearly the familiar semi-algebraic set.

Tarski was the first to show that every *Tarski set* can be represented by a quantifier free formula, and is therefore a semi-algebraic set. His algorithm however was complicated and many improvements were suggested, one of them, called *cylindrical algebraic decomposition* and suggested by Collins, is used here [C75].

Basically, Collins considers *standard prenex formulae* and the coresponding cells in the $\mathbb{R}^r$ space, which are bounded by zeros of the polynomials that specify the *atomic formulae* of the quantifier free part $\Phi(x_1, x_2, \ldots, x_r)$. He shows that the projection of such cells on the $\mathbb{R}^{r-1}$ subspace, is bounded by zeros of other polynomials in $r - 1$ variables. The number and degree of these polynomials are bounded. Collins shows that this projection leads to the elimination of the r-th variable. This procedure

may be extended to eliminate any number of quantifiers, leading to a quantifier free formula describing the cells of $\mathbb{R}^k$ for which the Tarski formula is true.

We are especially interested in the complexity of the cell arrangement in the reduced space. Let k_0 and m_0 be the number of polynomials defining the original quantifier free part of the formula, and their degree, respectively. After the quantifier elimination, we are left with semi-algebraic set specified by k_q polynomials of maximal degree m_q, where q is the number of quantifiers and

$$k_q \leq (k_0)^{2^q}(2m_0)^{3^{q+1}} \quad m_q \leq \frac{1}{2}(2m_0)^{2^q}. \tag{7}$$

A straighforward result of the Collins decomposition is that the extended boundary of a semi-algebraic set is also a semi-algebraic set. (For the proof and for more elaborated discussion of Collins decomposition, see [LB93].)

Using Collins results to find the structure of $K_{\bar{x}}^V$.

Collins results allow us also to find the structure of the parameter sets $K_{\bar{x}}^V$ that correspond to every point in the image plane.

Lemma 1 *For any semi algebraic set $V \subseteq \mathbb{R}^2$ of degree (k, m) $(m \geq 2)$ and for every $\bar{x} \in \mathbb{R}^2$, the set of transformation parameters*

$$K_{\bar{x}}^V = \{\bar{t} = (A', b') \,|\, \exists \bar{s} \;\; s.t. \;\; A'\bar{s} + b' \in V \;;\; ||\bar{x} - \bar{s}|| < \Delta \;\}$$

is also a semi-algebraic set of degree $(k_p = (k+1)^4(4m)^{27}, m_p = 0.5(4m)^4)$ (in the parameter space of affine transformations $\mathbb{R}^6$).

Consider now the parameter sets (in $\mathbb{R}^6$), associated with a set of points S. These parameter sets divide the parameter space into cells such that all parameter vectors in a certain cell induce the same membership relation: Let c be a cell in the parameter space $\mathbb{R}^6$ and assume that $t, t' \in c$. Then, $\forall x_i \in S$; $(V_t)^\Delta \cap x_i = (V_{t'})^\Delta \cap x_i$. This implies that the number of different $W(A, S)$ sets cannot exceed the number of cells in the parameter space $\mathbb{R}^6$ induced by the boundaries of $\{K_{\bar{x}}^V | x \in S\}$.

For N being the cardinality of S, the number of different subsets $A \subseteq S$ is 2^N, implying that the number of cells, or "connected components" in the parameter space is not lower. This however, is not possible for arbitrarily large points sets, because the intersection of polynomial (or semi-algebraic) sets cannot produce "too many" cells (Milnor theorem [M64]). This leads to the following relation which must be satisfied for every shattered set of cardinality N,

$$2^N \leq 2(2 + k_p m_p N)^6 \leq 2[2 + (k+1)^4(4m)^{31}N]^6 \tag{8}$$

The main theorem now follows by a straightforward calculation.

theorem 2 *For every semi algebraic open set V of degree (k, m) in $\mathbb{R}^2$,*

$$VCdim(C_{\mathcal{A}^\Delta}(V)) = O(\log km)$$

5 From localization to recognition

Consider a library $\mathcal{L} = \{V_1, V_2, \ldots, V_{\|\mathcal{L}\|}\}$ that contains several objects. According to the model based recognition paradigm, we assume that the object observed in the image is one of these objects after some transformation. Generalizing our predictions to the recognition case requires only to find the VC-dimension, associated with the hypotheses class $C_T(\mathcal{L}) = \bigcup_{i=1}^{\|\mathcal{L}\|} C_T(V_i)$.

theorem 3 *Consider the library of objects $\mathcal{L} = \{V_1, V_2, \ldots, V_{\|\mathcal{L}\|}\}$, and the class of transformations T. Let $C_T(V_i)$ be the class of concepts associated with this transformations class and with a particular object $V_i \in \mathcal{L}$. Let $C_T(\mathcal{L})$ be the concept class associated with the transformations class and with any of the objects in $\mathcal{L}$. Then,*

$$\frac{VCdim(C_T(\mathcal{L}))}{\log VCdim(C_T(\mathcal{L}))} \leq \log \|\mathcal{L}\| + \max_{i=1}^{\|\mathcal{L}\|} VCdim(C_T(V_i))$$

The theorem shows that the VC-dimension of a concept class that allows the concepts to come from different objects, grows slowly with the number of objects allowed, in a rate close to logarithmic function of the size of the library. This implies that the number of measurements sufficient to ensure that any hypotheses is close enough, also grows only by approximately a logarithmic factor. This is not surprising as it was observed experimentally that the computational effort involved in recognizing objects grows logarithmically with the number of objects considered [GTM89]. We have shown here that not only the computational effort but also the amount of information required to succeed in the task increases with logarithmic rate.

6 Conclusion

We analyzed the amount of data required to succeed in a recognition task under a particular probabilistic model, and obtained upper bounds on the number of data features required to draw a reliable hypothesis. The analysis was carried independently of the recognition method used, and in a certain sense, independently of the particular objects considered. The bounds predicts that more data may be required for more complicated objects, for tasks which require discrimination between similar instance, and for tasks which involve larger class of allowed hypotheses. They also predict that the amount of information required grows logarithmically with the objects complexity and the number of object in the library.

The results were obtained by showing the equivalence between recognition and a certain learning model, followed by using well-known techniques from learning theory. The key parameter that determines learning performance, and discriminate between difficult and easy tasks, was determined for the concept classes that correspond to various recognition tasks.

Describing recognition as a learning task throws a new light on common vision tasks such as reconstruction and model-based recognition, and demonstrates that these tasks are essentially equivalent, and differ only in the VC-dimension of the allowed hypotheses classes.

References

[BEHW89] Blumer, A., A. Ehrenfeucht, D. Haussler and M.K. Warmuth, 1989, "Learnability and The Vapnik-Chervonenkis Dimension", *JACM*, **36**(4), pp. 929-965.

[BL93] S. Ben-David and M. Lindenbaum, 1993, "Localization vs. Identification of Semi-Algebraic Sets", Proceedings of the 6th ACM Conference on Computational Learning Theory, pp. 327-336.

[C75] Collins, G.E., 1975, "Quantifier Elimination for Real Closed Fields by Cylindrical Algebraic Decomposition", Proceedings of the 2nd GI Conf. On Automata Theory and Formal Languages, *Springer Lec. Notes Comp. Sci.* **33**, pp. 515-532.

[GJ93] Goldberg P. and M. Jerrum, 1993, "Bounding the Vapnik-Chervonenkis Dimension of Concept Classes Parametrized by Real Numbers", Proceedings of the 6th ACM Conference on Computational Learning Theory, pp. 361-368.

[GTM89] Gottschalk P.G. , J.L. Turney, and T.N. Mudge, 1989, "Efficient Recognition of Partially Visible Objects Using a Logarithmic Complexity Matching Technique", *Int. J. of Rob. Res.*, **8**(6), pp. 110-131.

[GH91] Grimson, W.E.L., and D.P. Huttenlocher, 1991, "On the Verification of Hypothesized Matches in Model-Based Recognition", *IEEE Trans. on Pattern Analysis and Mach. Intel.*, **PAMI-13**(12), pp. 1201-1213.

[GHJ92] Grimson, W.E.L., D.P. Huttenlocher, and D.W. Jacobs, 1992, "A study of affine Matching with Bounded sensor error", *Second Europ. Conf. Comp. Vision*, pp. 291-306.

[L92] M. Lindenbaum, "Bounds on Shape Recognition Performance", 1993. Proceedings of the 7th ICIAP, Italy.

[LB93] M. Lindenbaum and S. Ben-David, "Applying VC-dimension Analysis to Object Recognition", 1993, CIS report No. 9330, Computer Science Department, Technion.

[M93] Maybank, S.J. ,1993, Probabilistic Analysis of the Application of the Cross Ratio to Model Based Vision", Manuscript.

[M64] Milnor, J., 1964, "On the Betti Numbers of Real Varieties", Proc. Amer. Math. Soc. **15**, pp. 275-280.

[MH90] Mundy, J.L. and A.J. Heller, 1990, "The Evolution and Testing of Model-Based Object Recognition Systems", Proc. 3rd ICCV, pp. 268-282.

[VC71] Vapnik, V.N. and A.Y. Chervonenkis, 1971, "On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities", Theory of Probability and its applications, 16(2), pp. 264-280.