Transforming an image into a data-flow for visual indexing

BOST, Jean Marc, MILANESE, Ruggero, PUN, Thierry

A recent trend in model-based object recognition is to build efficient systems for primary hypotheses generation. These so-called visual indexing systems, rely on the assumption that object identification can be performed by recovering local invariants. Our goal is to succeed in this task when objects to be identified are allowed to lie on an unknown background. As a unique restriction, we impose the background to be made of statistically homogeneous patterns.

The originality of our approach consists in transforming the initial static input image into a dynamic flow of data. To this end, the various primitives that are extracted from the image are temporally ranked. Ranks are computed so as to favor the most relevant features for visual indexing. From this stage, primitives are sequentially passed on toward the indexing process.

The architecture of our indexing system is composed of a set of hierarchical knowledge sources which update in parallel a blackboard structure. The asynchronism produced by the flow of input features is exploited by the activation strategy of knowledge sources. First deductions constrain future searches. In this way, a limited set of pertinent cues is efficiently recovered. Such cues are then globally interpreted so as to provide the most likely hypotheses.
1. Introduction

A recent trend in model-based object recognition is to build efficient systems, called visual indexing, for primary hypotheses generation. Up to date, several practical implementations have already provided valuable results [3,4,6,7,14].

In our project, we address the visual indexing task in a more general context. Our goal is to identify man-made objects lying on a non-uniform background made of repetitive patterns (Fig.1.a). Such context corresponds to that of many real applications, such as aerial imagery or industrial picking tasks. Moreover, we want the system to be able to perform on the sole basis of one unique input image. The approach we chose is inspired by a position paper from Rosenfeld [12] which basically states that local invariants are to be searched in parallel in order to provide cues for pruning the set of hypotheses.

Parallelism leaves us with at least one major problem. If we admit that all possible information is simultaneously available, then meaningful information is mixed with misleading information that is not related to the objects. In this paper, a solution is proposed by considering that all image information is not exactly available at the same time but is asynchronously processed by an asynchronous system. In this way, the first available information can be used to avoid future conflicting situations. Obviously, in this case the initial information must be selected so as to be meaningful for the objects to be identified.

2. Architecture for Asynchronous Visual Indexing

The idea of asynchronous processing is inspired by biological evidences and suggests to order input data after a delay that is proportional to their relevance [1]; one unique input static image is thus converted into a dynamic flow of data. The most important information for the task at hand is considered first and can be used to constrain next data.

A distributed architecture for asynchronous visual indexing, has been designed so as to process such a flow of data; a description of the architecture coupled with an attention system is provided in [9]. It is composed of a set of opportunistic knowledge sources (KS) which update in parallel a blackboard structure (BB) (Fig.2). The BB contains a hierarchy of perceptual groupings [8,3] and a solution island (SI). The SI contains the hypotheses that are compatible with the groupings. The KS’s function is to link elements of the BB in order to constitute higher order groupings. Each time a new grouping is built by a KS, some hypotheses are pruned out from the SI.

The basic element composing lowest-order groupings are called tokens. They are primitives that are extracted from the input image. In our case we use segments [2], arcs from contours [13], and intensity regions. Tokens of a same type are stored in a common structure called a token-map. Within such structures, tokens are temporally ranked as described below and, according to their rank, are sequentially passed on towards KS’s. Nevertheless, knowledge sources still acts in parallel on token-maps. Moreover, a KS can be implemented as a pipeline accepting a new token when the previous one is still being processed.
3. Relevance of tokens

Fig. 1.d illustrates how segments from contours have been ranked in our specific context. It can be noticed that the brighter segments are globally rather related with scene objects than with segmentation artifacts or background patterns. Consequently, this makes them more suitable for interpretation. Fig. 1.b and Fig. 1.c show the two criteria that are combined in order to obtain such ranking.

The first criterion, called reliability, represents the likelihood of a token to have a physical correspondence in the scene rather than to be a segmentation artifact. Consider, for example, the segment-map. The longer and sharper a segment is, the more likely it is to correspond to the linear boundary of a real object [5]. Table 1 contains a list of attributes that have been selected to evaluate the reliability of our tokens. Let \( m \) be a token from a token-map \( M \), let \( A_r(M) \) be the set of attributes for judging the reliability of tokens in \( M \) and let \( \bar{a} \) be the normalized value of attribute \( a \); the reliability, \( r \), of \( m \) is given by:

\[
r(m) = \sum_{a \in A_r(M)} \bar{a}(m)
\]

A reliable token can however be issued from a foreground object as well as from background patterns. This is the reason why another measure, called significance, is introduced. Such measure indicates how likely a token is to belong to a target object.

Techniques for the detection of conspicuous features in an image have been successfully applied in our project to focus the attention on an object [10]. The same techniques are also useful to distinguish object tokens from background tokens. Basically, a token is significant if some of its attributes make it different from the majority of the other tokens within the token-map. The attributes that have been selected in our experiments for segments, arcs and regions, are summarized in Table 1. Let \( A_s(M) \) be the set of attributes for evaluating the significance in \( M \). The significance of a token \( m \in M \) is given by:

\[
s(m) = \sum_{a \in A_s(M)} c(m, a)
\]

where \( c(m, a) \) is the conspicuity of \( m \) for attribute \( a \) and indicates the percentage of tokens within the token-map that have a different value for attribute \( a \) than \( a(m) \).

4. Time precedence of tokens

In order to determine the temporal rank of tokens, reliability and of significance are combined. We define the relative precedence \( \rho \), of a simple token \( m \in M \) as:

\[
\rho(m) = \bar{r}(m) \cdot \bar{s}(m) \quad (\sim \text{’s indicate normalizations})
\]

Such combination provides the rank of a token with respect to other tokens in the same token-map. However, relative precedences obtained on two different token-maps are computed on a different number of attributes which themselves cannot be compared. Consequently, one
token-map $M$, may have a much higher average value of $\rho$ than other token-maps. This would cause the indexing process to start analyzing tokens almost exclusively in $M$. This would bias the identification process, waste the possibility of parallelism and reduce the robustness of the process.

In order to have a robust identification which does not rely on a single token-map, the probabilities of tokens to be observed first should be balanced over all token-maps. Let $m_1 \in M_1$ and $m_2 \in M_2$ be two tokens from different token-maps, let $\pi(m_1)$ and $\pi(m_2)$ be their ranks; the a-priori probabilities of $m_1$ and $m_2$ to have a given rank must be equals:

$$\forall x \in [0,1] \quad P(\pi(m_1) = x) = P(\pi(m_2) = x) \quad (4)$$

In order to satisfy Eq.4 the a-priori probabilities must be estimated. For each token-map, we estimate their distribution from a set of training sample. Such estimation can then be transformed into a new distribution by equalization of the histogram [11]. For example, the estimated distribution of a token relative precedence $\rho$, from a sample of precedences $\{\rho_1,...,\rho_n\}$, can be converted into a uniform distribution by means of a function $S$:

$$S(\rho) = \sum_{\rho = 1}^{\rho_n} [(\rho - \rho_{j-1}) \cdot n_j + (\rho - \rho_n) \cdot n_{k+1}]$$

where $k$ is defined such that $\rho_{k+1} > \rho \geq \rho_k > \rho_{k-1} > ... > \rho_1 > \rho_0 = 0$, and where $n_j$ is the number of occurrences of $\rho_j$. $S(\rho)$ can be proved to verify Eq.4. Finally, given $S$, the precedence $\pi$, of an observed token $m \in M$ is given by:

$$\pi(m) = S(\rho(m)) \quad (6)$$

5. Activation Strategy for KS’s

An asynchronous prototype for visual indexing is currently being implemented. During the identification process, the KS activation is guided by a control strategy which exploits the delays between times of token availability. Indeed, delays in the recovery of high-level groupings are used to update the SI. As an example, we report in Table 3 the average precedence for a pair of parallel segments when target objects are isolated. It can be noticed that for linear objects such as the pen or the stapler, the average precedence value is higher than for circular objects like the cup or the adhesive band. Hence, during the identification process a high precedence value for the detection of parallel segments is used to reject the hypotheses “cup” and “adhesive band” from the SI. More generally, all groupings produced by the KS’s have a precedence value from which conclusions about the hypotheses to be kept in the SI are drawn.

If, after analysis of a few tokens, the SI already contains a reliable solution, then remaining tokens are only considered when they are consistent with such solution. Consequently, all possible groupings are not included in the SI and conflicting situations do not arise.
6. Conclusion

Experiments on the use of temporal precedences in visual indexing are ongoing. First results are promising and confirm the importance of the concept. Other applications for visual indexing of delays between availability time of tokens are now under study.

References:

Table 1: Selection of attributes for reliability and significance

<table>
<thead>
<tr>
<th>Token-map</th>
<th>Selected attributes for reliability</th>
<th>Selected attributes for significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segments</td>
<td>expected length, contrast</td>
<td>length, orientation</td>
</tr>
<tr>
<td>Arcs</td>
<td>expected radius, length, contrast, fit quality</td>
<td>radius, turning angle, length</td>
</tr>
<tr>
<td>Regions</td>
<td>expected size, contrast, shortness of boundary</td>
<td>size, intensity</td>
</tr>
</tbody>
</table>

Table 2: Average precedence of first parallel segments for 4 typical objects

<table>
<thead>
<tr>
<th>object</th>
<th>cup</th>
<th>band</th>
<th>pen</th>
<th>stapler</th>
</tr>
</thead>
<tbody>
<tr>
<td>average precedence</td>
<td>0.79</td>
<td>0.67</td>
<td>0.97</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Figure 1: Reliability, significance and relative precedence of segments

Figure 2: General architecture for visual indexing