At the Interface Between Vision and Language

Daniel Chester

Vision + Language

- parsing charts
- solving puzzles

Vision

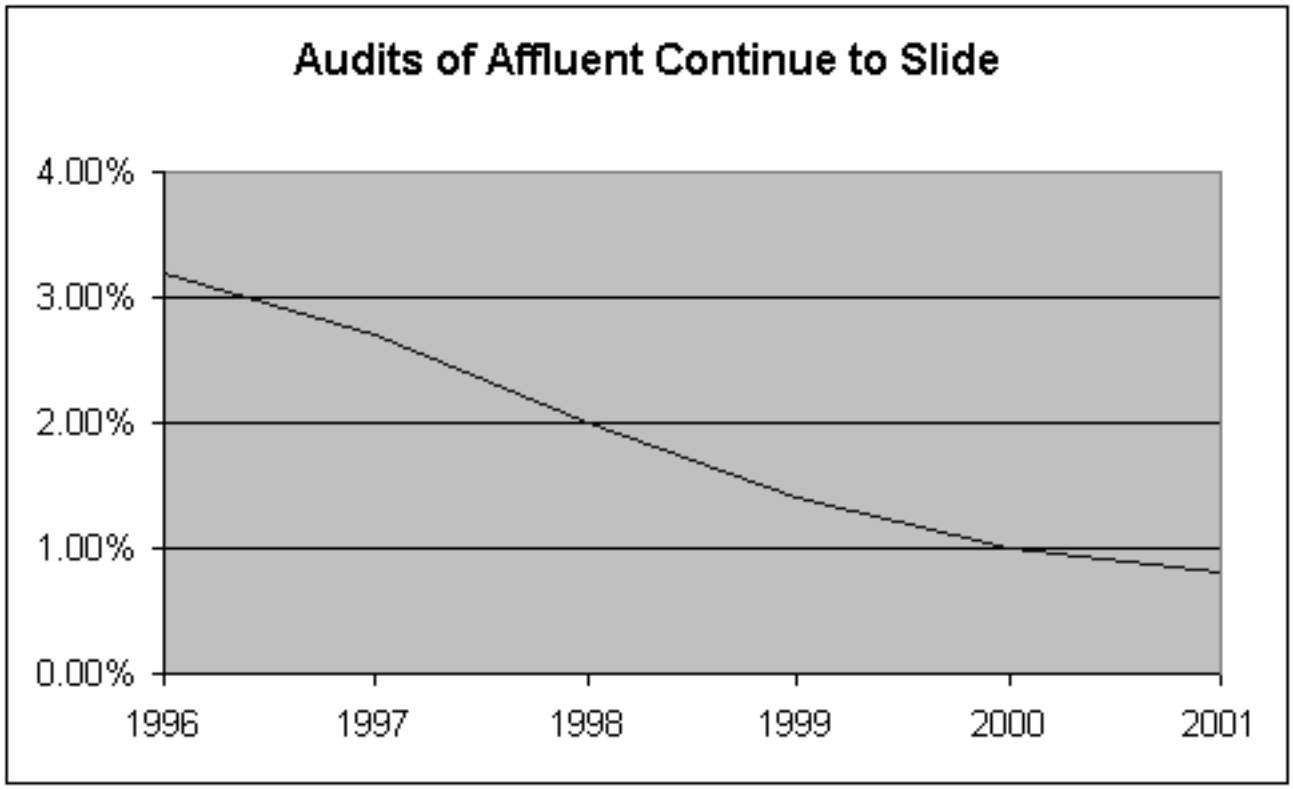
- detecting edges

Language

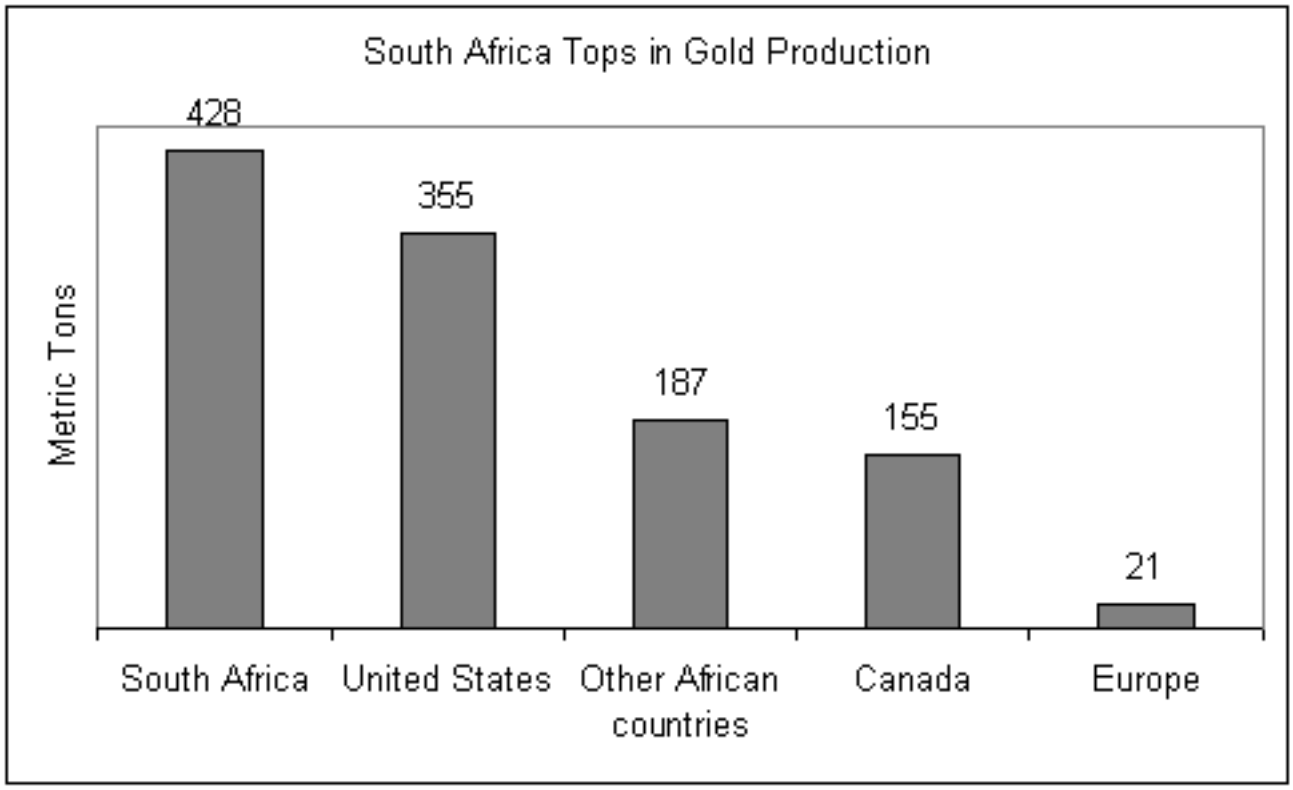
- parsing pragmatically

Parsing Bar and Line Charts

The extraction of information from bar and line chart images requires the recognition and interpretation of lines, curves, geometric shapes, words and phrases used in titles and labels, and recognition of common conventions used in charts.


Audits of Affluent Continue to Slide
4.00%
3.00%
2.00%
1.00%
0.00%
1996
1997
1998
1999
2000
2001

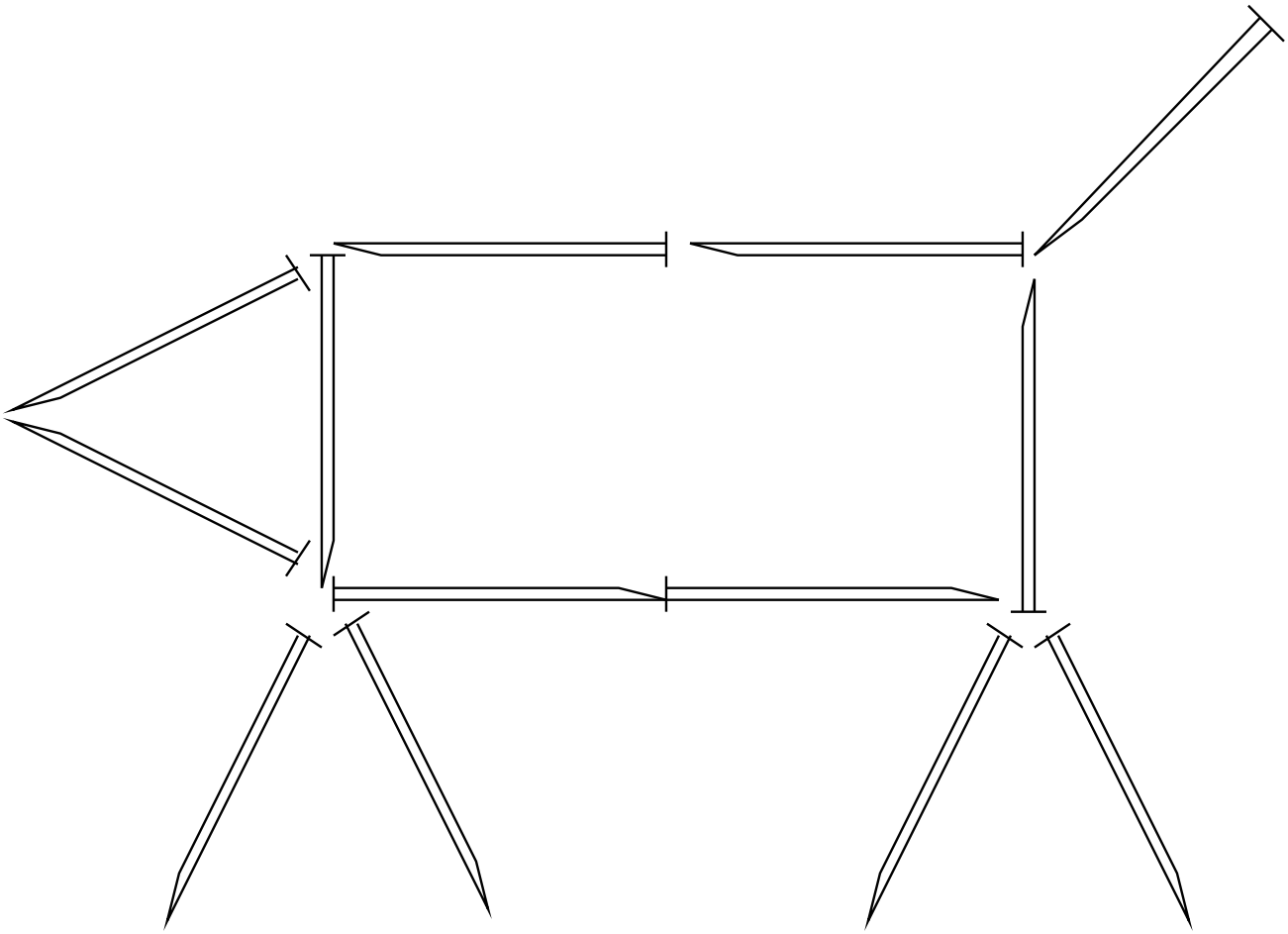
```
<InformationGraphic Caption=
 "Audits of Affluent Continue to Slide">
<LineChart>
<XAxis>
<Label></Label>
</XAxis>
<YAxis>
<Label></Label>
</YAxis>
<line>
<point>
<X>1996</X><Y>3.191</Y>
</point>
<point>
<X>1997</X><Y>2.702</Y>
</point>
<point>
<X>1998</X><Y>2</Y>
```


South Africa Tops in Gold Production
Metric Tons
428
355
187
155
21
South Africa
United States
Other African countries
Canada
Europe

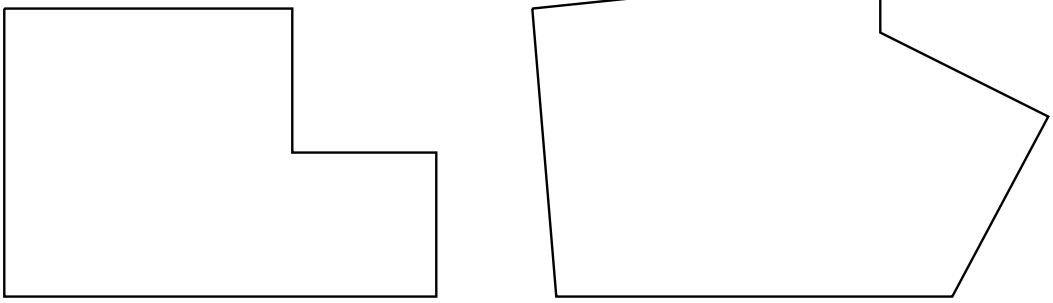
```
<InformationGraphic Caption=
 "South Africa Tops in Gold Production">
<BarChart BarDirection="vertical">
<MeasurementAxis>
<Label>Metric Tons</Label>
</MeasurementAxis>
<BarAxis>
<Label></Label>
</BarAxis>
<Bar>
<Label>South Africa</Label>
<Color>128</Color>
<Height>178</Height>
<Annotation><Value>428</Value></Annotation>
</Bar>
<Bar>
<Label>United States</Label>
<Color>128</Color>
```

Solving Word + Picture Problems

Word + picture problems such as are found in puzzle books are solved with a minimum of preprocessing by humans.



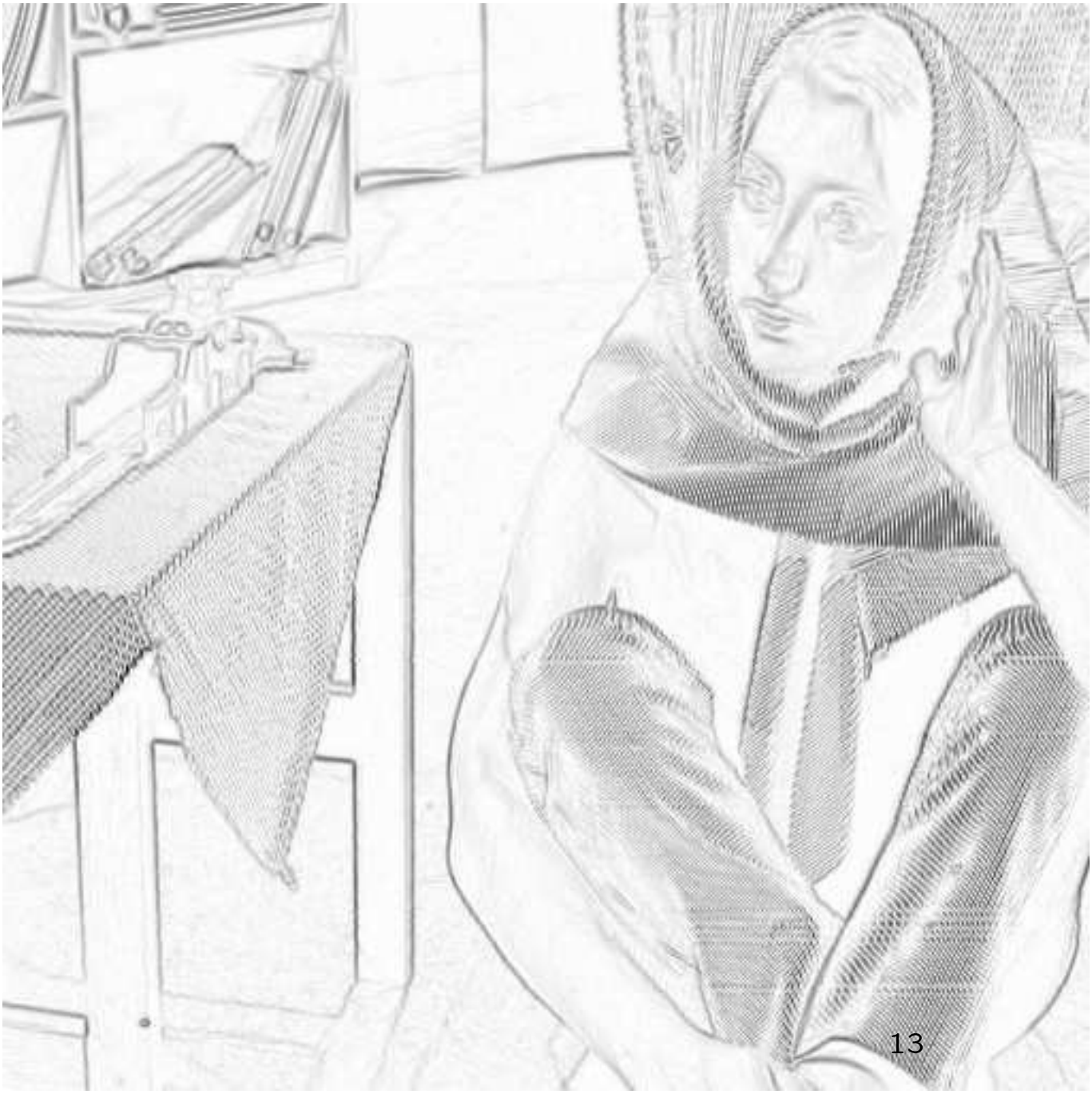
If you move just two nails, can you make this dog face the other way - and still keep his tail in the air?



Each of these shapes can be cut into three pieces, which can then be rearranged to form perfect squares. Where would you make the cuts?

Edge Detectors that Preserve Fine Detail

Most edge detectors blur images and lose fine detail. We have a pair of edge detectors that preserve fine detail.



A More Pragmatic Approach to Parsing

A look at the wide range of natural languages and of the many ungrammatical ways that humans use to communicate suggests that language parsing/understanding should be more pragmatic than the context-free-grammar based approaches now being used.