

AI Agent to Solve 2x1 and 2x2 Visual Analogy Problems in Image Format

Magahet Mendiola
(mmendiola3@mail.gatech.edu)

March 19th, 2015

Contents

Introduction	1
Visual Approach	1
Visual Heuristic and Answer Selection	2

describing the way your algorithm works, its relative strengths, its relative weaknesses, its efficiency, its relationship with human cognition, and the different challenges and opportunities in solving problems visually rather than verbally.

In addition to your agent, you should also write and submit a design report of roughly 1500 words. This design report should do a number of things. First, it should describe the reasoning your agent uses. How does it work? Second, it should describe how the agent comes to some of its correct answers. Third, it should describe why your agent makes some of the mistakes it does. Fourth, it should describe what could be done to improve the agent if you had more time, resources, or processing power. Fifth, it should describe the unique challenges and opportunities in visual reasoning compared to verbal reasoning. Your design report can include diagrams in addition to the ~1500 words.

Introduction

We will describe the design of an AI agent built to solve 2x1 and 2x2 visual analogy problems given images rather than object descriptions. This will include details on the agent's strengths, weaknesses, efficiency, and relationship with human cognition.

Visual Approach

Given visual input rather than descriptive data provides both challenges and opportunities for our AI agent. On one hand, propositional reasoning is made more difficult as we would first have to translate pixel-based information into a descriptive model of object types, sizes, and inter-relationships.

On the other hand, visual input allows us to broaden our agent's approach to understanding the inter-figure relationships and simplifies the application of transformations that should be applied to the entire figure. For example, figure 1 shows a difficult problem to solve with propositional reasoning. First, we would have to reason that the rotation of the single triangle in frames A and B should be applied to all three triangles in C. Second, we would need to model how rotating all three triangles would affect their spatial relationships. Third, we would have to address the correspondence problem between comparing triangles in frame C and the answer frames.

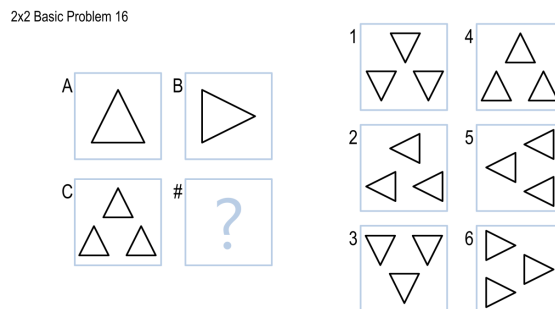


Figure 1: 2x2BasicProblem16

These complexities are avoided altogether with a visual approach. Once we observe that the transformation from A to B is a 90° rotation, it is trivial to apply the same rotation to frame C and compare the result to the answer choices. There is no need to teach the agent the concepts of triangles, changes in spatial relationships, or object correspondence.

The visual approach, for this example, seems to match much more closely to human cognition. We observe the rotation in the example frames and apply it, effortlessly, to frame C. We do not need to enumerate the spatial relationship dynamics between the three triangles in order to rotate the image.

This example illustrates the comparative strengths of the visual approach to solving RPMs. In the

following sections, we will examine a visual heuristic algorithm that solves better than half of the provided RPM test cases with simplistic reasoning and no prior understanding of shapes, spatial relationships, or transformation models.

Visual Heuristic and Answer Selection

Our agent performs the following steps for each problem:

1. Load each figure as bilevel (black and white) images.
2. Create a vector for each image of the black pixel count in each quadrant (top-left, top-right, bottom-left, bottom-right).
3. Get the absolute pairwise difference between example source and destination frame vectors.
4. Get the absolute pairwise difference between the target and each answer choice frame vectors.
5. Select the answer frame with the smallest euclidean distance (l-2 norm) between the example and candidate transition vectors.

This