A NEW PROOF OF THE NP-COMPLETENESS OF VISUAL MATCH

R.A. Rensink

Technical Report 89-22
Sept 1989

A NEW PROOF OF THE NP-COMPLETENESS OF VISUAL MATCH

R.A. Rensink

Technical Report 89-22
Sept 1989

Department of Computer Science
The University of British Columbia
Vancouver, BC, Canada V6T 1W5

Abstract

A new proof is presented of Tsotsos' result [1] that the VISUAL MATCH problem is NP-complete when no (high-level) constraints are imposed on the search space. Like the proof given by Tsotsos, it is based on the polynomial reduction of the NP-complete problem KNAPSACK [2] to VISUAL MATCH. Tsotsos' proof, however, involves limited-precision real numbers, which introduces an extra degree of complexity to his treatment. The reduction of KNAPSACK to VISUAL MATCH presented here makes no use of limited-precision numbers, leading to a simpler and more direct proof of the result.

A New Proof of the NP-Completeness of Visual Match

1 Introduction

The problem of VISUAL MATCH was formulated by Tsotsos [1] as a way of demonstrating the inherent intractability of most (if not all) bottom-up visual processing. He started from the observation that visual search is a component of many low-level visual processes, be they instantiated in man or machine. By reducing visual search to the problem of VISUAL MATCH, and showing VISUAL MATCH to be NP-complete, he demonstrated that visual search is NP-complete when no general constraints are imposed on the search process. Since NP-complete problems are generally assumed to be intractable, simple use of any realistic amount of parallel computation will not ensure the rapid execution of bottom-up visual search. Additional constraints and/or top-down control are also required.

The proof given by Tsotsos [1] is based on the polynomial reduction of the NP-complete problem KNAPSACK [2] to VISUAL MATCH. Tsotsos' proof, however, involves limited-precision real numbers, which introduces an extra degree of complexity to his treatment. The reduction of KNAPSACK to VISUAL MATCH presented here makes no use of limited-precision numbers. This leads to a simpler and more direct proof that makes clear the essential similarity of the two problems.

1.1 The VISUAL MATCH problem

The essence of the VISUAL MATCH problem is to determine whether a particular pattern (the *goal*) is present in a given image (the *test image*). Since a match is unlikely to be perfect, it becomes necessary to determine whether there is some subset of the test image which matches the pattern sufficiently well. The following discussion (based on [1]) formulates this problem more precisely.

Let $\mathbf{I}$ denote the test image, a set of pixels indexed by the spatial coordinates $(x, y) \equiv \mathbf{x}$. Let the set of coordinates be denoted $\mathbf{X}$. The value of $\mathbf{I}$ at any given point $\mathbf{x} \in \mathbf{X}$ is an m-dimensional vector $\mathbf{i}(\mathbf{x})$, corresponding to a set of m different measurements. Each measurement corresponds to a scene parameter, and is represented as an integer [1]. For notational convenience, the jth component of $\mathbf{i}(\mathbf{x})$ will be denoted as $i(x, y, j)$.

The goal image $\mathbf{G}$ is similarily defined. Measurements need not be exactly of the same types as those in $\mathbf{I}$. Measurement types common to both test and goal images are assigned the same index j. Different types in $\mathbf{I}$ and $\mathbf{G}$ can be accomodated by expanding the dimensionality of the vectors appropriately; if a measurement type does not exist in the test or goal image, assigned it a zero value. Let M denote the dimensionality of the resultant vectors.

The requirement that test and goal images correlate to an acceptable degree is given by

$$\sum_{\mathbf{x} \in \mathbf{X}'} \mathbf{corr}(\mathbf{x}) = \sum_{\mathbf{x} \in \mathbf{X}'} \sum_{j=1}^{M} g(x, y, j) \times i(x, y, j) \geq \kappa,$$

where $\mathbf{X}'$ is some subset of locations common to the test and goal images, and $\kappa \in Z^+$ is the correlation threshold. Note that summation over noncontiguous locations is allowed. Intuitively, this provides for the exclusion of points where noise or the interposition of other objects have altered the measurements(s) in the test image.

Since any superset of the goal image will lead to a high degree of "correlation", it is also necessary to penalize differences. This is done by imposing the constraint

$$\sum_{\mathbf{x} \in \mathbf{X}'} \mathbf{diff}(\mathbf{x}) = \sum_{\mathbf{x} \in \mathbf{X}'} \sum_{j=1}^{M} |g(x, y, j) - i(x, y, j)| \leq \beta,$$

where $\mathbf{X}'$ is the same subset of locations as above, and $\beta \in Z^+$ governs how much difference is acceptable.

Since the elements of $\mathbf{I}$ and $\mathbf{G}$ only influence the correlation and difference functions, they are not really necessary for any once these two functions have been determined for any particular instance. As such, VISUAL MATCH can be described as:

Instance: A finite set $\mathbf{X}$; a "correlation" $\mathbf{corr}(\mathbf{x}) \in Z^+$ and "difference" $\mathbf{diff}(\mathbf{x}) \in Z^+$ for each $\mathbf{x} \in \mathbf{X}$; a correlation goal $\kappa \in Z^+$, and a difference constraint $\beta \in Z^+$.

Question: Is there a subset $\mathbf{X}' \subseteq \mathbf{X}$ such that

$$\sum_{\mathbf{x} \in \mathbf{X}'} \mathbf{corr}(\mathbf{x}) \geq \kappa \quad \text{and} \quad \sum_{\mathbf{x} \in \mathbf{X}'} \mathbf{diff}(\mathbf{x}) \leq \beta \ ?$$

[1] Measurements can be represented as fixed-precision real numbers, as done in [1], but multiplication by an appropriate factor can transform them into integers.

2 The NP-Completeness of VISUAL MATCH

Like most proofs of NP-completeness, VISUAL MATCH will be demonstrated to be NP-complete by showing (i) that it is a member of the class NP, and (ii) that it is NP-hard (i.e., that a known NPC problem will reduce to it in polynomial time). Taken together, these two conditions are sufficient to ensure that VISUAL MATCH is NP-complete [2].

i) VISUAL MATCH $\in$ NP

Proof: Both the **diff** and **corr** functions are computed by simple operations at each point, followed by summation over the relevant measurement types. As such, computing their value at any point clearly requires only polynomial time. Given any test image **I'**, the relevant pointwise measurements can be simply added together, and their values tested against the specified thresholds. Thus, VISUAL MATCH is clearly in NP. ■

ii) VISUAL MATCH is NP-hard

Proof: Consider first KNAPSACK, an NP-complete problem described as follows [2]:

> **Instance:** A finite set U; a "value" function $v(u) \in Z^+$ and "size" function $s(u) \in Z^+$ for each $u \in U$; a value goal $K \in Z^+$, and a size constraint $B \in Z^+$.
>
> **Question:** Is there a subset $U' \subseteq U$ such that
>
> $$\sum_{u \in U'} v(u) \geq K \quad \text{and} \quad \sum_{u \in U'} s(u) \leq B \ ?$$

Given an instance of KNAPSACK, let v_{max} denote the maximum value of $v(u)$, and s_{max} the maximum value of $s(u)$. Now, consider an instance of VISUAL MATCH defined in the following way. Let the measurement space of the test and goal images be the same, and set its dimensionality be $v_{max} + s_{max}$. Next, set both test and goal images to be arrays one pixel high and $|U|$ pixels wide. (The shape of the images does not affect bottom-up visual match.) Associate each element of U with a particular location in this array.

Once the assignments have been made, the values of the measurements in the test and goal image can be specified. Given any element $a \in U$, assign to the corresponding location $\mathbf{x}_a \equiv (x_a, 1)$ the following values:

test image: set the first $v(a)$ elements of $\mathbf{i}(\mathbf{x}_a)$ to 1, and the next $(v_{max} - v(a))$ elements to 0. Set the next $s(a)$ elements to 1, and the remaining elements to 0.

goal image: set the first v_{max} elements of $\mathbf{g}(\mathbf{x}_a)$ to the same values as the corresponding elements of $\mathbf{i}(\mathbf{x}_a)$. Set the remaining s_{max} elements to 0.

Notice now that for any element a,

$$\mathbf{corr}(\mathbf{x}_a) = \sum_j g(x_a, 1, j) \times i(x_a, 1, j) = v(a), \quad (1)$$

and

$$\mathbf{diff}(\mathbf{x}_a) = \sum_j |g(x_a, 1, j) - i(x_a, 1, j)| = s(a). \quad (2)$$

so that the selection of elements in KNAPSACK corresponds exactly to the selection of locations in VISUAL MATCH. To complete the proof, it suffices to note that all operations in the reduction can clearly be carried out in polynomial time. ■

References

[1] Tsotsos JK, *The Complexity of Perceptual Search Tasks*, Technical Report RBCV-TR-89-28, Department of Computer Science, University of Toronto, April 1989.

[2] Garey MR, and Johnson DS, *Computers and Intractability: A Guide to the theory of NP-Completeness*, New York: W.H. Freeman & Co., 1979