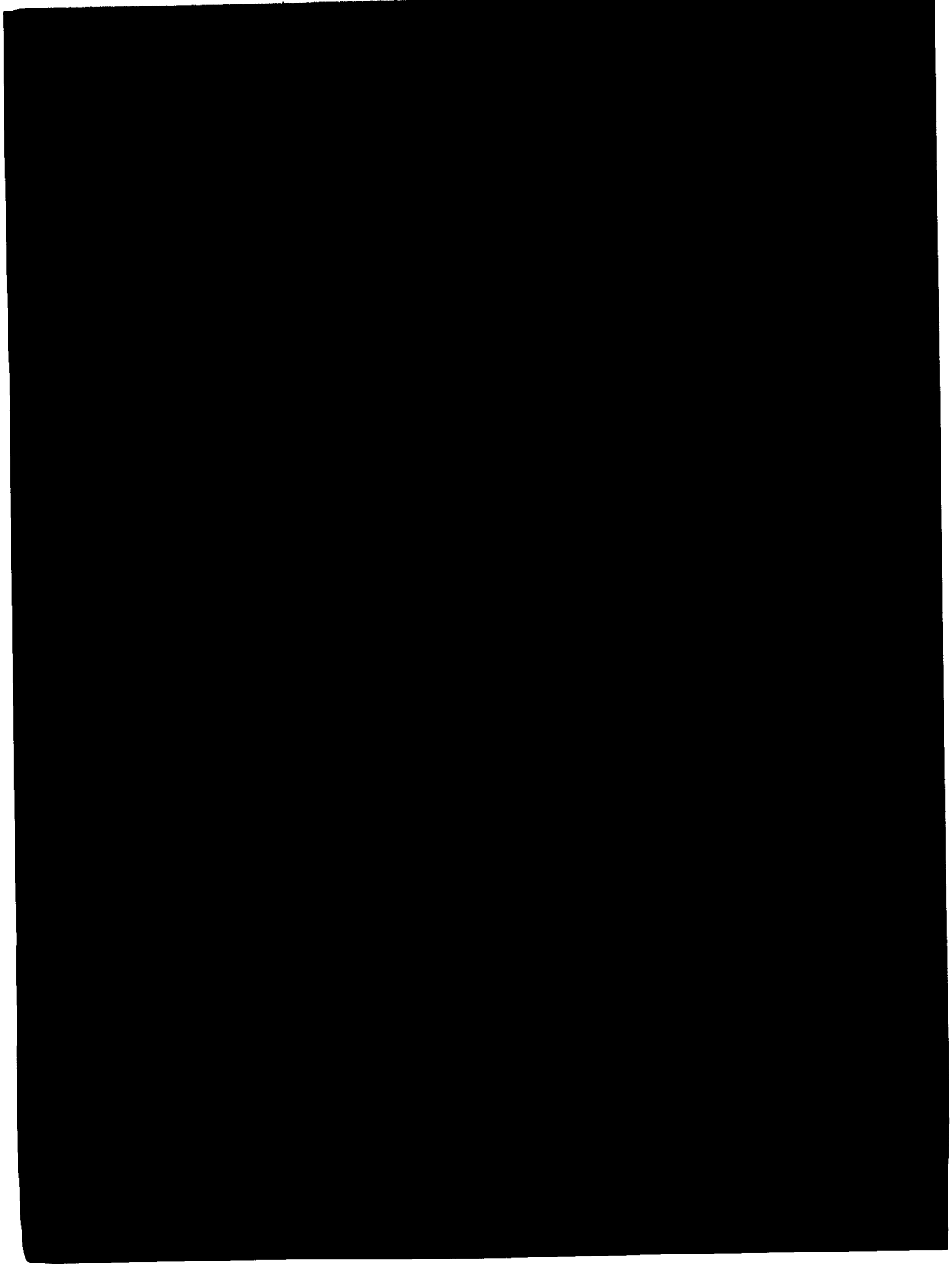

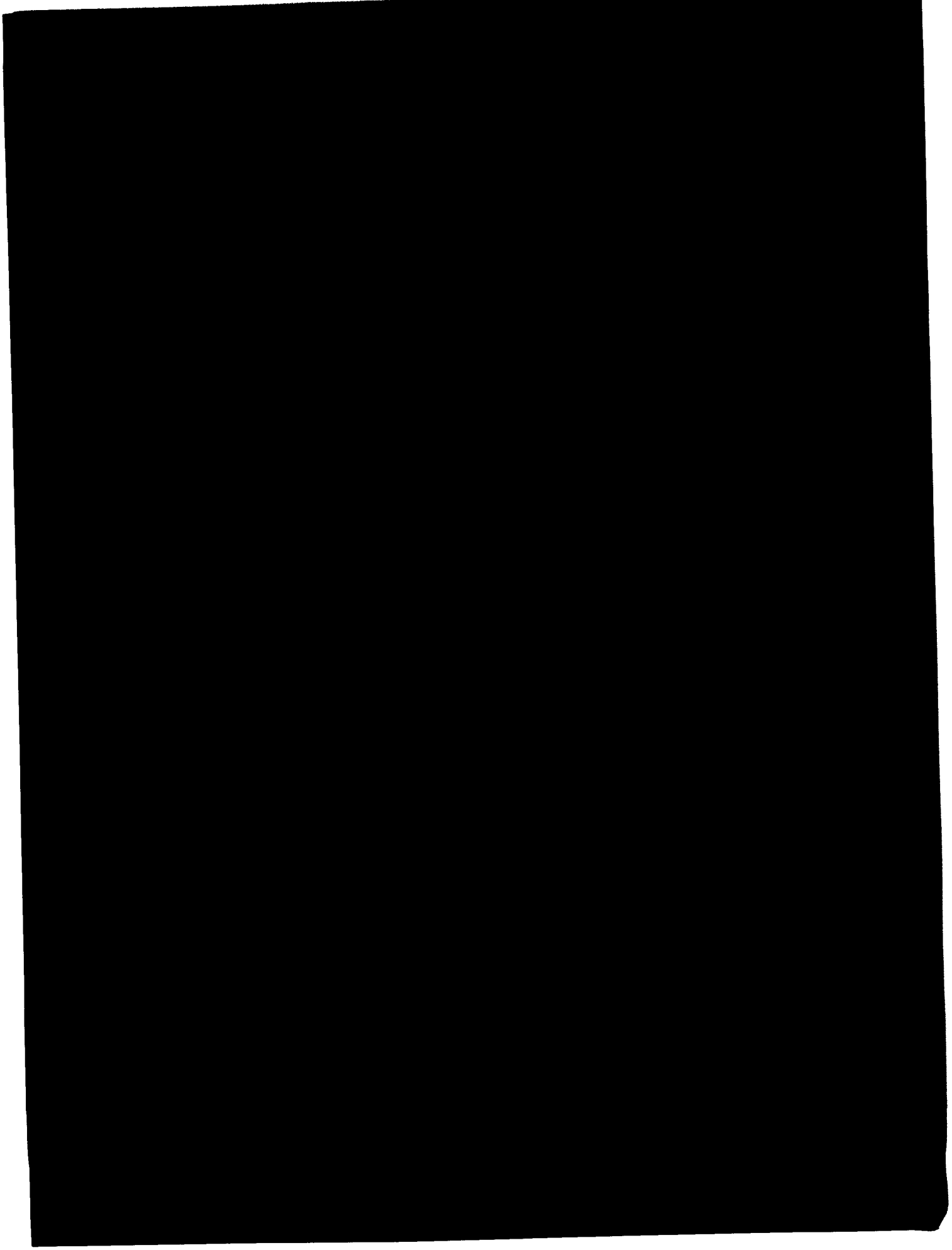

Artificial Intelligence 59 (1993) 95–101
Elsevier



ARTINT 1002

# From a real chair to a negative chair

Takeo Kanade

*School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA*

## 1. How it started

In the fall of 1977, I was invited by Raj Reddy to spend a year at the Computer Science Department of Carnegie Mellon University as a visiting scientist from Kyoto University in Japan. My plan for research during the stay was to develop a model-based object recognition program. Upon arrival, I chose an image of an office scene (Fig. 1) as an example image; the image was one of a set that Ron Ohlander had used in his research on color image segmentation. The task I set for my program was to recognize the chair in this image.

I began to write a "knowledge-based" program for chair recognition by creating a set of heuristic rules for the task. It seemed that in addition to geometric relationships, a good source of constraints was color information, such as "the back and the seat of a chair have the same color". The effort of creating heuristic rules one after another, however, was not a satisfying game, since every time I came up with a reasonably functioning program, I could also find a chair that was an exception to the rules.

## 2. The Origami world

Ohlander's color segmentation program, which was quite famous at the time, could segment the image of Fig. 1 into regions as shown in Fig. 2. As I stared at the results, not only could I still see a chair without using color information, but I could also perceive the shape of the object without

*Correspondence to*: T. Kanade, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA. E-mail: kanade@cs.cmu.edu.

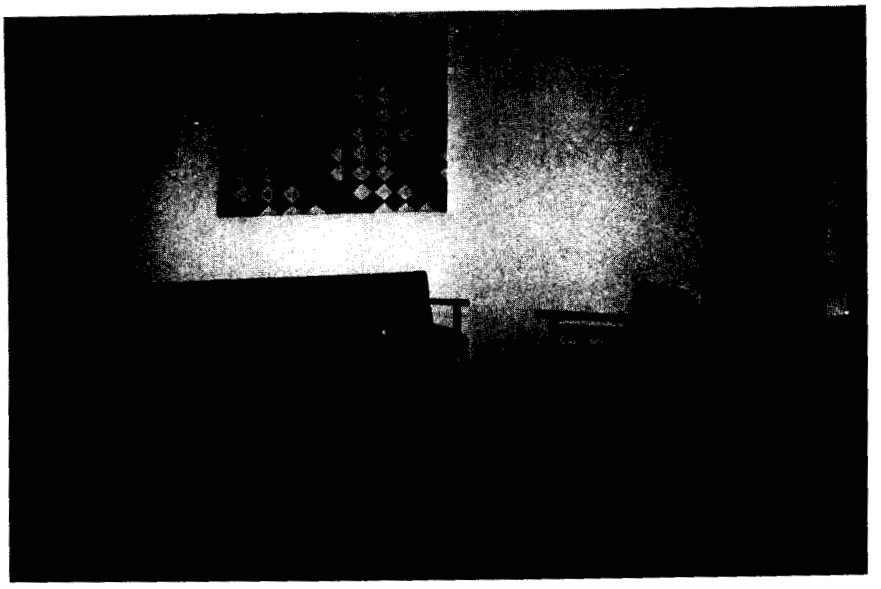

Fig. 1. An office scene—one of Ohlander's image set.

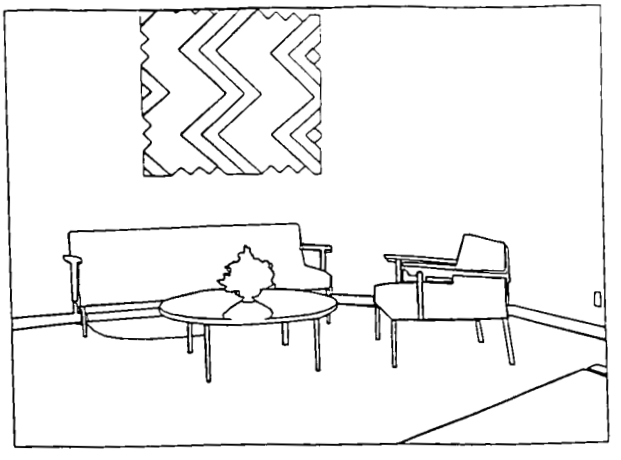

Fig. 2. Ohlander's segmentation result.

knowing it was a chair. In other words, I realized that there must exist geometrical constraints which enable us to recover shape from a single image.

To make the problem simpler, I turned Fig. 2 into a more stylized line drawing of a chair as shown in Fig. 3, and again started to develop heuristic rules for shape recovery, this time using geometric constraints only. Being heuristic, this effort did not lead me to a satisfactory systematic solution,

either. Six months had passed since I arrived at CMU, and I was beginning to feel frustrated.

One day, Allen Newell asked me to come to his office and talk about my research. I felt embarrassed about not having achieved much yet. When I finished explaining what I had been doing, Newell asked what the difference is between my method and Waltz's line labeling method [6]. I gave a very vague answer, "Waltz's method cannot handle this image, and mine will be more flexible". When I returned to my office, I quickly found the exact reason why Waltz's labeling method did not work with Fig. 3. Its trihedral world assumption (i.e., at each vertex three planes meet) implies solid objects, and while the chair is a solid object, pictures like Fig. 3 typically fail to depict the width of thin objects, such as legs, and thus become "illegal" drawings. The two drawings of a box in Fig. 4 are good illustrations of the effect.

So, to generalize beyond the trihedral world, I constructed a shape labeling theory for the world consisting of planar surfaces which may be folded, cut, or glued together only along straight lines. From the metaphor of the Japanese art of paper folding, I named it a theory of the Origami world [1].

## 3. Multiple interpretations and qualitative shapes

The labeling procedure for the Origami world could interpret line drawings like Fig. 3 and Fig. 4(b) as well as Fig. 4(a), but revealed two interesting

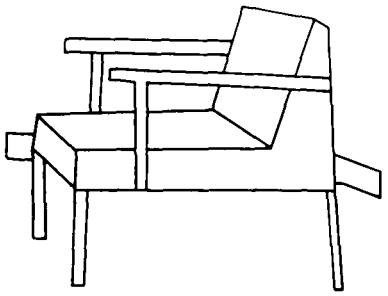

Fig. 3. A "chair" line drawing.


Fig. 4. (a) A box in the trihedral world and (b) a box in the Origami world.

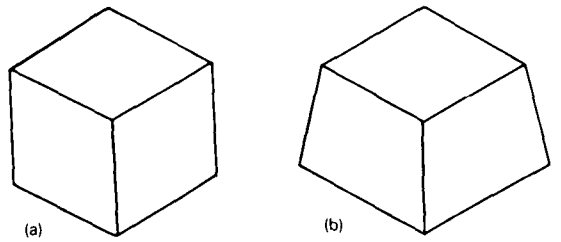


Fig. 5. (a) A "cube" line drawing and (b) a "trapezoid" line drawing.

issues, both of which are obvious upon afterthought. The first phenomenon was that it generated many legal interpretations. For example, Fig. 4(b) was given eight interpretations. For the chair scene in Fig. 3, my program generated several hundred interpretations after spending two hours of a DEC-10's CPU time, but it was far less than halfway through enumerating all the possible interpretations. Though many of these interpretations appear strange to us, all of them are correct and *physically possible* shapes that can correspond to the drawings. I used to say, "If you bet that this drawing is of a chair, your bet is one to at least several hundred or against two CPU hours of DEC-10, so most likely you will lose". [1]

Generating multiple interpretations for a single drawing is certainly counter-intuitive, since humans don't usually do so. Mathematically, however, it is obvious that a drawing cannot determine a unique shape, since a 2D picture is only a projection of a 3D scene, and many 3D shapes can produce the same projection. The trihedral world of Waltz and others was so constrained that it had not needed to face multiple interpretations often. The Origami world is bigger than the trihedral world, and thus began to reveal the issue.

The second issue that the Origami world revealed was the realization that the labeling does not really specify the shape. Till then, people talked as if the Waltz labeling procedure had recovered the shape. Actually, line labels can only analyze and recover a qualitative characterization of the shape. This is clear from Figs. 5(a) and 5(b): each figure can have the same labeling or qualitative shape (convex corner), but the figures suggest different quantitative shapes (cube and trapezoid). While the POLY system of Mackworth [5] used the gradient space as a quantitative tool, it did so only for the purpose of *analyzing* the feasibility of shapes, not for *recovering* the quantitative shape.

[1] When I told this to my current students, their response was "DEC-10 must have been a slow machine".

## 4. Quantitative recovery of plausible shapes

It became clear that we cannot avoid multiple interpretations, and that we cannot recover "the" shape that a drawing depicts. Still, however, the problem remained as to why certain shape interpretations appear more plausible than others. To say "a certain shape is more familiar" or "I learned it over the years", though not incorrect, simply ducks the question and does not really answer it, since most of us, in fact, *cannot* think of multiple interpretations. We do not select a particular interpretation after we think of all the possibilities; rather we think of *only* a certain shape. Thus, additional shape constraints must be used at a relatively early stage of the interpretation process.

My goal became the development of a quantitative geometrical theory to shed light on this problem. The two drawings in Fig. 5 turned out to be the simplest example that illustrated the problem: they have exactly the same set of three qualitatively different interpretations, they each have only three quadrilateral regions one of which (the top region) is exactly the same in each, but the seemingly most plausible quantitative interpretations are different, i.e., cube versus trapezoid. Thus, the two side regions must be responsible for the difference.

The idea I came up with was a principle of non-accidental regularity. That is, regular image properties, such as collinearity, parallelism, and regular spacing, occur for a good reason, not by accident. John Kender, who was working on his thesis on texture, phrased it as the "seeming so is actually so" principle [4]. I devised a method to convert the non-accidental regularities into shape constraints.

One of the most interesting new classes of the non-accidental regularity I defined was skew symmetry. Real symmetry in 2D has an axis for which the opposite sides are reflective; in other words the symmetry is found along lines perpendicular to the symmetrical axis. The concept of skew symmetry relaxes this condition, such that symmetry is found along lines not necessarily perpendicular to the axis, but at a fixed angle to it. Figures like a leaf and a parallel quadrilateral have skew symmetry. Non-accidentalness of skew symmetry implies "A skewed symmetry depicts a real symmetry viewed from some (unknown) viewing direction" [2, p. 424]. This results in quantitative constraints on surface orientations that include a skew symmetry, and the constraints can be represented very neatly in gradient space. Using the similar idea, I found that various regular image properties, including parallelism, regular spacing, and affine-transformable texture patterns, can be mapped to constraints on plausible shapes.

Once we combine these quantitative constraints with the qualitative shape characterizations given by labeling, we can compute the particular plausible shapes that a drawing can depict. The principle of non-accidental regularity

and its mathematical embodiment in the skew symmetry heuristic, together with a method to compute the quantitative shape, constituted the main part of the *AI Journal* paper that is included in the set of the most influential works [2].

Using the theory, it turned out that the "normal" interpretations of many drawings (a box for Fig. 4(b), a cube for Fig. 5(a) and a trapezoid for Fig. 5(b), respectively) could be proven to be the most plausible. For the chair scene of Fig. 3, however, my program found two "most plausible" interpretations. One of them was a "normal" chair, and the other was a strange shape which I called a "negative" chair.

Although it looks odd, a negative chair is physically realizable, satisfies the non-accidental regularity principle, and, of course, its picture can look the same as the normal one. Interestingly, when we construct a negative chair and swing it a little in the air, it appears as if a normal chair is flexing its arms back and forth. This effect awaits a psychophysical explanation.

## 5. Computer vision as a physical science

One of the main contributions of the paper "Recovery of the three-dimensional shape of an object from a single view" is that it demonstrated a simple fact in vision: there are a multiplicity of possible image interpretations, and if we want to obtain a unique interpretation, we must use some assumptions, constraints, or heuristics. Since humans usually think of only a single interpretation, many vision researchers accepted the requirement that a computer vision program also generate only a single interpretation. In fact, when I submitted the Origami world paper [1] to the *Artificial Intelligence Journal*, one of the reviewers gave a comment like, "This paper is wrong since it does not give the interpretation 'box' to the drawing of a box". Early researchers attempted to meet this requirement by hastily incorporating domain heuristics, often, implicitly, without understanding their effects, limitations, or implications. In contrast, the paper "Recovery of the three-dimensional shape of an object from a single view" showed how far one could go purely with geometrical principles, and what exact specification was for the set of possible interpretations by that method.

A series of works appeared in the last decade which formalized many of the geometrical constraints which relate properties in the image domain to three-dimensional shape constraints. Some of these constraints may have been conceived from observations of human perception, and they may be heuristic in the sense that they do not always hold. The implications were clearly defined, however, and therefore it was possible to predict the consequences when rules did not apply. In this sense they are not *ad hoc*. This is in direct contrast to the purely heuristic methods, ranging from

various line drawing interpretation methods in the early 1970s, to the use of global minimization of arbitrarily created energy functions in the 1980s. In these cases, the implications are neither clearly defined nor predictable in terms of physical reality.

Systematic formulation of constraints in vision need not be limited to geometric constraints. Recent attention has turned to putting other physical properties, such as the optical and statistical processes which underlie vision, into a quantitative, computational framework. That is, the emphasis is on developing *physical* models for computer vision. Vision is the process of using images (data) taken about the real (physical) world. Therefore, as with any physical science, a clear technical understanding of the physical nature of the data (i.e., images) is required for formulating a solution. Such modeling reveals the structure of visual information: the exact information that is contained in an image, the limits of processing algorithms, and the heuristic knowledge required to resolve any remaining ambiguity. Hence, I advocate computer vision is a physical science as well as a cognitive science [3].

## References

[1] T. Kanade, A theory of Origami world, *Artif. Intell.* **13** (1980) 279–311.

[2] T. Kanade, Recovery of the three-dimensional shape of an object from a single view, *Artif. Intell.* **17** (1981) 409–460.

[3] T. Kanade, Computer vision as a physical science, in: R. Rashid, ed., *Carnegie Mellon Computer Science: A 25-Year Commemorative* (Addison-Wesley, Reading, MA, 1991) 345–369.

[4] J.R. Kender, Shape from texture, Ph.D. Thesis, Carnegie Mellon University, Pittsburgh, PA (1980).

[5] A.K. Mackworth, Interpreting pictures of polyhedral scenes, *Artif. Intell.* **4** (2) (1973) 121–137.

[6] D. Waltz, Generating semantic descriptions from drawings of scenes with shadows, in: P.H. Winston, ed., *The Psychology of Computer Vision* (McGraw-Hill, New York, 1975).