A New Paradigm For Testing Human
And Machine Motion Perception

Thomas V. Papathomas, Andrei Gorea
A new paradigm for testing human and machine motion perception

Thomas V. Papathomas* and Andrei Gorea†

*AT&T Bell Laboratories, Visual Perception Research Department, 600 Mountain Avenue, Murray Hill, NJ 07974
†Université René Descartes and CNRS, Laboratoire de Psychologie Expérimentale 28 rue Serpente, 75006 Paris, France

ABSTRACT

We present a new paradigm for studying motion perception. This approach is based on a class of stimuli that we devised for testing the relative strength of stimulus attributes (luminance, color, spatial frequency, orientation, binocular disparity, etc.) in eliciting motion perception by correspondence matching. In this class of stimuli, different attributes are matched simultaneously in the spatio-temporal domain in a systematic, algorithmic manner which allows each attribute to produce motion in an arbitrary direction (if, of course, it is a token for movement perception), independently of the other attributes. This results in animation sequences in which many different motion paths may co-exist, each path due to a different attribute. Such an arrangement allows a direct comparison of the strength of attributes in eliciting movement. Results from psychophysical experiments based on our paradigm can be used to develop complex motion detection models for machine vision systems which attempt to approximate human performance. Similar methods are discussed for studying stereopsis mechanisms.

1. INTRODUCTION

One of the problems in the study of motion is finding the relative strength with which the various object attributes (luminance, color, spatial frequency, shape, etc.) elicit movement perception. In continuous, but particularly in stroboscopic (apparent), motion the term correspondence matching refers to the process by which the visual system manages to match the attributes of moving objects across space and time to detect any motion of the targets. Thus the question is whether we can design experiments which will enable us to classify these attributes according to their importance in this spatio-temporal correspondence matching. For example, is luminance more important than shape in eliciting motion perception? How does spatial frequency rate when compared to, say, orientation?

Significant progress has been made in answering questions of this type. We would like to begin, albeit anachronistically, with a special kind of stimulus, used, among others, by Navon1 and by M. Green and his colleagues2,3 which enables the experimenter to examine whether the studied attribute is a token in movement perception and, if so, how it compares with other attributes. The idea behind Green's stimulus is illustrated in Figure 1a, in which the "targets" are shown as circles. Solid and dashed circles indicate what is displayed in frames 0 and 1, respectively.

All the attributes of the targets are kept the same, except for attribute A, which can be assigned N different values, 0, 1, ..., N-1 (in Figure 1a N = 2). As can be seen from the spatio-temporal arrangement of Figure 1a, if attribute A is indeed a token for the perception of movement, the sequential

Figure 1. (a) An example of the stimuli used by M. Green and his collaborators. The objects drawn in solid and dashed lines are displayed in frames 0 and 1, respectively. (b) A similar arrangement where the targets are all on the same line. The same convention is used as in (a) with respect to solid and dashed circles.
display of frames 0 and 1 will give rise to motion in the counter-clockwise direction, as indicated by the path $p_R$, because the inter-frame displacement $\Delta \theta$ is half the period $P_R$. Using stimuli of this type Navon\(^1\) observed that shape is not a strong token for carrying movement. Green and Odom\(^2\) showed that the third dimension, i.e. depth, is indeed important in the perception of motion. Green\(^3\) demonstrated that orientation and spatial frequency are also good tokens for motion, whereas phase is not. What Green was not able to do was to compare directly the strength of the various tokens in eliciting apparent motion, although he mentions that “orientation also proved to be a correspondence token, although anecdotal observer reports suggest that it is not as powerful a token as spatial frequency.”\(^4\) We will describe, in Section 3, an extension of this stimulus which offers a way to compare any pair of attributes directly.

Other researchers have used similar stimuli, the most common of which is the type shown in Figure 1b, where the elements are configured along a line and where the motion is along the horizontal direction around the circle $p^R$. To generalize, we have shown an inter-frame displacement $\Delta x$ which is not necessarily half the spatial period $P_x$, as is the case in Figure 1a (of course, this could also be done with the stimulus of Figure 1a). Experiments have been conducted in which the strength of motion $S$ was recorded as a function of $\Delta x$. One may compare indirectly the strength of various attributes in eliciting motion by comparing the corresponding curves of the functions $S(\Delta x)$ for those attributes. Again, our paradigm offers a more direct method of comparing the strength of two attributes.

2. The Role of Various Attributes in Motion

Stimuli similar to those presented in Section 1, as well as numerous variations of those, have been used by experimenters to investigate the role of the various attributes in eliciting movement perception. We shall mention briefly the work of several researchers who have contributed in this area, concentrating on results that are relevant to our paradigm.

In addition to Navon\(^1\) and Green\(^2\) who experimented with shape in general, and orientation in particular, Kolvers and Pomerantz\(^2\) and Kolers\(^5\) presented evidence indicating that these attributes are very weak tokens for carrying motion. Similar conclusions were drawn by Hochberg and Brooks\(^6\), Watson\(^7\) and Burt and Sperling\(^8\). On the other hand, Ullman\(^9\), Green\(^10\), and Gorea and Papathomas\(^11\) report that orientation can indeed carry motion, albeit not as strongly as spatial frequency can. The apparent disagreement between these two groups can be easily explained by the different spatio-temporal arrangements and the different criteria employed. The attribute that was studied most extensively is luminance. Several researchers have experimented to investigate its role in motion and we limit our list to a few recent reports by Antiss\(^12\), Cavanagh et al.\(^13,14\), Gorea and Papathomas\(^15\), and Chang and Julesz\(^16\). There is general agreement that luminance if one of the, if not the, strongest carriers of movement. Another strong carrier is spatial frequency, as the results of Gorea and Papathomas\(^16\), Green\(^2\) and Ramachandran, Ginsburg and Antiss\(^17\) indicate. The roles of size and contrast were investigated by Ullman\(^9\) and Pantle and Sekuler\(^18\), respectively. Binocular disparity proved important in motion (Green and Odom\(^3\), Papathomas et al.\(^19\)). As far as color is concerned, all studies agree that it is a much weaker token than luminance (Cavanagh et al.\(^13,14,21\), Troscianko\(^22\), Ramachandran and Gregory\(^23\)); some experimenters have carried the argument further by theorizing that color is not at all a carrier of motion (Livingstone and Hubel\(^24\)).

3. A New Paradigm

In all of the stimuli used in the apparent motion experiments mentioned earlier, all the attributes of the objects-targets were held fixed except for the one attribute that was being studied. There are a few exceptions to this observation, but even in the few cases where two attributes were varying simultaneously (Ramachandran et al.\(^17\), Cavanagh et al.\(^1\) and Kooi et al.\(^25\)) the methods used do not allow the extension of their paradigm to any number of arbitrarily chosen attributes, as is the case with our new class of stimuli which allows any multiple attributes to be arranged simultaneously in the spatio-temporal plane.

3.1 Correspondence Matching with Multiple Attributes — A New Class of Stimuli

The best way to introduce this class of stimuli for motion perception is to describe it in general terms and then to illustrate it with specific examples. The general idea in the case of clockwise/counter-clockwise motion along a circle is shown in Figure 2. To understand the notation used in this Figure, which shows schematically four time frames $t = 0, 1, 2$ and $3$, the reader is asked to compare it to Figure 1a, which shows schematically only two frames.

For simplicity’s sake let us initially consider only the first two frames of Figure 2, i.e. $t = 0$ and 1. Although the radius of $t = 0$ is shown larger than that of $t = 0$ for notational convenience, in fact the elements of $t = 0$ and $t = 1$ are displayed in temporal sequence on the same fixed radius $r_0$, just as in Figure 1a. At $t = 2$, the elements of the third ring $(t = 2)$ are displayed at the same radius $r = r_0$, followed by the elements of the outer ring $(t = 3)$ again at $r = r_0$. The only reason we increase the radius of the second and subsequent frames is to show the spatio-temporal patterns that give rise to clock- or counter-clockwise motion. Thus, the increasing radius in Figure 2 corresponds to increasing time in a manner that is directly analogous to the Cartesian $x - t$ plane of Adelson and Bergen\(^26\) in horizontal motion. By analogy, Figure 2 introduces the $\theta$-plane, i.e. the polar form of the $x - t$ plane; of course, as we indicated elsewhere\(^11\), our technique can also be applied in the $x - \theta$ plane.

Each target in the stimulus of Figure 2 can be characterized by many attributes $A, B, C, \ldots$ Only two of them are shown here for simplicity (we shall see how these attributes can be combined effectively in subsection 3.2). Each attribute $X$ can take on $N$ different values $X_0, X_1, \ldots, X_{N-1}$ ($N = 3$ in Figure 2). Other choices of assigning values to $X$ will be considered at the end of this subsection. For example, if $A$ is color, then one possible assignment would be $A_0 = \text{red}, A_1 = \text{green}, A_2 = \text{blue}$; if $B$ is orientation, then $B_0, B_1$ and $B_2$ could correspond to $0^\circ$, $120^\circ$ and $240^\circ$, respectively, and so on. In our stimuli, we usually choose the inter-frame displacement $\Delta \theta$ to be half the inter-element distance $r_0$, but this is not necessary. Next let us consider the types of motion resulting from the configuration of Figure 2. If we concentrate on the variations of attribute $A$, let us begin with the element $A_2B_2$ of $t = 0$ at the $5$-o’clock position. At $t = 1$ it is more likely that, as far as attribute $A$ is concerned, that element will be matched to the $A_2B_0$ element of $t = 1$ at $4$-o’clock rather than to the $A_2B_3$ element of $t = 1$ at $6$-o’clock. Proceeding in this manner, we see that, if attribute $A$ is indeed a token for motion, this element will elicit a counter-clockwise motion along the circle of radius $r_0$, indicated by the path $p_{AB}$. Similarly, if we concentrate on attribute $B$, we will observe a clockwise path $p_{BA}$ in the opposite direction to that of $p_{AB}$.

Before considering other general properties of this kind of stimuli, let us illustrate with a simple example. Figure 3 illustrates a specific instance of Figure 2, where $A$ and $B$ are spatial frequency and orientation, respectively. In this case, $A_0$, $A_1$, and $A_2$ are low, medium, and high spatial frequency and $B_0$, $B_1$, and $B_2$ are assigned orientations of $0^\circ$, $60^\circ$ and $120^\circ$.

Note that, if you “follow” the spatial frequency (ignoring the orientation changes), the resulting path is in the counter-clockwise direction as $t$ increases. The dual is true if you follow the orientation without paying attention to the changes in spatial
frequency: the resulting trajectory is in the opposite (clockwise) direction.

One of the properties of these stimuli is that they are periodic in time (see subsection 3.3.3). Here it can be verified that the temporal period is 6 frames. Thus we would need to generate two more frames (t = 4 and t = 5) and then we could create an animation sequence of arbitrarily long duration. When we display these frames in sequences of long duration, and show each frame sufficiently long to allow smooth pursuit of the targets, the resulting percept is quite interesting. If one concentrates in spatial frequency and "locks his/her attention" to, say, the low frequency A0, one may see these "balls" move in the counter-clockwise direction while, at the same time, they rotate about their center (also in the counter-clockwise direction) in increments of 60°. If one "locks in" to a fixed orientation, say horizontal, one may see these balls move in the clockwise direction while their stripes become thicker or thinner.

If, on the other hand, we show sequences of only a few (say 2 to 4) "short-living" frames (typically 33 msec), to prevent eye movements, then the observer presumably cannot attend to a particular attribute and perceives the motion in the direction that is dictated by the strongest between the two attributes. This allows the researcher to design experiments for the direct comparison of motion "tokens".

Of course, there is no reason why we could not arrange both attributes A and B in the same (concurrent matching) rather than in opposite (competitive matching) directions on the spatio-temporal θ-τ plane. If both of them are tokens for motion, such an arrangement would enhance the strength of the elicited apparent movement. Indeed, we verified this statement for many pairs of attributes such as binocular disparity and orientation, or luminance and color, etc.

Another issue that must be emphasized is that this paradigm does not limit us to two attributes. In fact, an arbitrary number of attributes can be combined as long as it is possible to form a conjunction of multiple attributes on the same target simultaneously (we present our ideas on that in the next subsection). Suppose it is possible to use many attributes A, B, C, ..., Z. The power behind our paradigm is that each of these can be arranged in the spatio-temporal domain independently of the others. A subset of them may be arranged to produce motion in one direction, while another subset may be configured to elicit movement in the opposite direction, while a third subset may be arranged to produce ambiguous motion, i.e. motion that is equally probable to be seen in two opposite directions. There are at least three ways to produce ambiguous
Figure 3. A particular example of Figure 2, in which $A$ is spatial frequency and $B$ is orientation.

Figure 4. Illustrating the cyclic configuration of attribute $C$ in the $\theta-t$ domain.
motion using, say, attribute C. For all three we must have $\Delta \theta = \frac{1}{2} P \theta$. 1) The simplest one is to let C stay fixed at $C_0$ for all the targets. Then, as far as attribute C is concerned, clockwise motion is equally probable as counter-clockwise motion. 2) The same is true when C is distributed cyclically, varying as shown in Figure 4.

The reason the name cyclic was chosen is that, whether you follow the clockwise path $p_F$ or the counter-clockwise path $p_R$, the values of C change in a cyclic fashion: $C_0, C_1, C_2, C_0, C_1, C_2, \ldots$ for $p_F$ and $C_2, C_1, C_0, C_2, C_1, C_0, \ldots$ for $p_R$. Again, C is not matched along either path, resulting in ambiguous motion. 3) Finally, another way to elicit ambiguous motion is to assign to the targets random values of C from a set of possible values $\{C_0, C_1, C_2, \ldots, C_n\}$.

Most of the studies of other researchers concentrate on studying one attribute A, while keeping all others $(B, C, \ldots, Z)$ fixed. This is what we call motion due to A within B, C, \ldots, Z. If, on the other hand, we configure attribute A to carry motion while we vary attribute C either cyclically or randomly, then we call this kind of motion as occurring across attribute C. This is another advantage of our paradigm, i.e. the ability to observe different modes of movement.

3.2 Conjunctions of attributes

The success of designing stimuli of the type described in the previous subsection depends on our ability to form conjunctions of many different attributes within the same target. To fix matters, let us consider the following attributes: $A =$ spatial frequency, $B =$ orientation, $C =$ color, $D =$ luminance, $E =$ contrast, $F =$ size (scale), $G =$ binocular disparity. In general, then, a target may be located at $(x_0, y_0)$ and be assigned a value $A_iB_jC_kD_tE_rF_sG_p$ where each attribute $X$ is present in the stimulus and is assigned a value $X_q$, where $X \in \{A, B, C, D, E, F, G\}$ and $0 \leq q \leq N-1$. One possibility of a target which may combine these attributes is the Gabor patch 27, shown in Figure 5.

![Figure 5. A Gabor patch.](image)

Its luminance function $I(x, y)$ is given by

$$I(x, y) = I_0 + J \exp[(x^2 + y^2)/\sigma^2] \sin[\omega_x x + \omega_y y + \phi]$$

with

$$x_* = x - x_0 \quad \text{and} \quad y_* = y - y_0,$$

where $I_0 =$ mean luminance, and $J$ is the (luminance) amplitude. The fundamental spatial frequency and the orientation of the patch are functions of $\omega_x$ and $\omega_y$, and $\phi$ determines the phase. Such a patch may be designed to possess unique values of $A$, the fundamental spatial frequency; $B$, orientation; $C$, color; $E$, contrast; $F$, size and $G$, binocular disparity. It is not clear how to assign values to the luminance attribute in this case, although one possibility is to manipulate $I_0$. Nevertheless, this element offers the possibility of forming conjunctions of a large number of attributes.

3.3 Properties of the new class of stimuli

These stimuli possess some desirable properties11 which we shall present briefly:

3.3.1 Ease of generation. Let us denote by $f$, $r$ and $c$ the subscripts assigned to attribute $X$ at location $\theta$ and $t$ (see Figure 2) so that the resulting arrangement will produce clockwise, counter-clockwise and cyclically ambiguous motion (see Figure 4), respectively. Then, there exist explicit formulas for these subscripts:

$$f = (\theta + t) \% (2N)/2$$

$$r = ((\theta - t) \% (2N))/2$$

$$c = (-\theta) \% N$$

where a $b$ is the remainder of the division of a by b, mod is the modulo function, and $N$ is the number of different values that are assigned to each attribute. The spatiotemporal positions occupied by targets are only those for which $\theta + t$ is even. Hence the values of $f$, $r$ and $c$ are only defined for $(\theta + t)$ even.

3.3.2 Periodicity along $\theta$. The elements are periodic along $\theta$ with period $N$, as can be verified either from Figure 2 or, in general, from Eq. (3).

3.3.3 Periodicity along t. Even more important than spatial periodicity, the stimuli possess temporal periodicity with a period $P = 2N$. This allows us to create animation sequences of arbitrarily long duration by displaying frames 0 through $2N - 1$ over and over again.

4. OTHER APPLICATIONS OF THE PARADIGM

In addition to using these stimuli for motion experiments with human subjects, there are other areas where such stimuli can find applications.

4.1 Correspondence matching in stereopsis

The Cartesian form of the stimuli can be used to test the relative strength of two attributes for stereopsis. The idea is illustrated in Figure 6, which is drawn after a similar figure by Julesz28, used to illustrate the construction of a classical random-dot stereogram (RDS). The extension that we propose here is that, instead of using elements characterized by one attribute only (typically luminance in classical RDSs), we use the conjunction of multiple attributes.

The symbols $A$, $a$ and $\alpha$ all refer to the same attribute, say luminance, which will be denoted generically by $\bar{A}$. We show two values of luminance: $A_1 = \alpha_1 = \beta_1 = \text{bright}$ and $A_0 = \alpha_0 = \text{dim}$. Similarly, $B$, $b$ and $\beta$ all stand for a second attribute, say color, denoted by $\bar{B}$. One possible assignment is $B_0 = b_0 = \beta_0 = \text{blue}$ and $B_1 = b_1 = \beta_1 = \text{yellow}$.

The reason why we use three different symbols for the same attribute will be explained next. For simplicity, ignore the presence of the symbols $B$, $b$ and $\beta$ in Figure 6, for the time being. With respect to $\bar{A}$, one may distinguish three different areas in the left and right images: 1) There is an area that is identical in both images and it forms the background; the symbol $A$ is used in this area. 2) There is a central square $(4 \times 4)$ outlined by dashed (left) or dashed/dotted (right) lines, which is also identical in the two images, but it is shifted horizontally to the right by one block in the left image relative to the right image. This forms the target and the symbol $a$ is used in this area. 3) Finally, because of the shift, there are areas that become uncovered and have no correspondence in the two images. The symbol $\alpha$ is used in these areas. When fused in stereo, the target is seen to loom over the background. This is the method of constructing classical RDSs.28

We now present our extension of the RDS by considering attribute $\bar{B}$ in Figure 6. Here, if we ignore the presence of attribute $\bar{A}$, we also observe a background (symbol $B$), a square target, outlined by dotted (left) or dashed/dotted (right) lines
Figure 6. Using conjunctions of attributes to generate random-dot stereograms.

(symbol b), and the uncovered areas (symbol β). Here, however, the target is shifted horizontally in the opposite direction, resulting in a stereoscopic percept of a square "hole" dug behind the background. The question becomes: When each element (block) is defined by the conjunction of attributes $A_iB_j$, as shown in Figure 6, what will be the resulting percept? How will the two attributes interact? We are planning experiments to address such questions.

Of course, there are other configurations one can construct. One possibility is to use more than just two values for each attribute. Another variant is the case in which $A_i$ may be distributed as shown in Figure 6, while $B_j$ is assigned random values; such a pairing could be used to study the strength of $A_i$ across $B_j$ in stereopsis. Yet another possibility is to form a looming (i.e. cross-disparity) square using $A_i$ and a looming triangle using $B_j$ and observing which of the two figures dominates, thus obtaining a measure of the relative strength of the two attributes for stereopsis.

4.2 Tests for motion detectors

We are currently completing a set of experiments that we have originated for assessing the relative strength of several attributes for motion detection. In particular, we have used the competitive matching paradigm (subsection 3.1) with several pairs of attributes to determine which one is stronger. One interesting result is that, when we showed such stimuli with long frame durations and long sequences where the competing attributes were color and orientation, some subjects saw "layered" motion; i.e. they reported that they perceived simultaneously two "layers" sliding across one another in opposite directions. It would be interesting to input these stimuli to various motion detection models and compare their performance to that of humans.

5. MOTION DETECTION MODELS

Several computational models have been proposed recently for animal and human motion perception. A pioneer in this area is the model of Reichardt29 on which is based the ERD (elaborated Reichardt detector) model of van Santen and Sperling30. The ERD model has been shown to be equivalent to that of Watson and Atumade31, as well as to the energy-based model of Adelson and Bergen32. Some of these can be modified to serve as motion detectors in machine vision. Most of these models, however, sense motion that results from spatio-temporal variations only in the luminance values of the stimuli. As a result, the proposed model detectors are assumed to be narrowly tuned to the Fourier energy of the stimulus at various spatial and temporal frequencies (although Chubb and Sperling33 have recently developed a more general model which detects motion in "non-Fourier", drift-balanced stimuli).

Our studies indicate that other important attributes must be incorporated in such motion perception models. For example, color seems to be a carrier of motion, albeit weaker than luminance. Binocular disparity was also found to play a role in recovering three-dimensional representations for the purposes of motion. The above comments are valid for both long-range and short-range motion mechanisms12,33,34,35 but are particularly valid for the low-level, short-range models. As far as the long-range mechanisms are concerned, which are thought to play a role in the correspondence matching process, our observations on the complex interactions of several attributes can be used to fine-tune the performance of motion models for better approximating human performance.

6. CONCLUSIONS

This new paradigm for studying motion perception allows the simultaneous and independent arrangement of multiple attributes in the spatio-temporal domain. The ensuing animations involve very complex motion paths and enable us to study the interactions of the attributes in the process of producing apparent motion. These stimuli allow us to study directly the relative strength of various carriers of motion. At the same time such stimuli and ideas can also be used for testing animal motion models of machine motion detectors. Finally, similar ideas, based on the conjunction of attributes, can be used in the study of stereopsis.

7. ACKNOWLEDGEMENTS

The authors wish to thank Ted Adelson, Pat Cavanagh, Ji-Jie Chang, and John Krauskopf for helpful discussions. The ideas of subsection 4.1 were developed jointly with Bela Julesz, whose help and encouragement for the entire project are greatly appreciated.
8. REFERENCES


35. O. Braddick, "Low-level and high-level processes in apparent motion," Philosophical Transaction of the Royal Society of London B290, 137-151 (1980).