References

Blackman, S. and Popoli, R. (1999) *Design and Analysis of Modern Tracking Systems,* Artech House, Norwood, MA, pp 1004-1018.

Bolderheij, F. and Van Genderen, P. (2004) Mission Driven Sensor Management.: *Proc. 7th Int. Conf. on Information Fusion*, Stockholm, pp 799-804

Bolderheij, F., Absil, F.G.J. and Van Genderen, P. (2005) Risk-Based Object-Oriented Sensor Management, *Proc. 8th Int. Conf. on Information Fusion*. Philadelphia.

Bolderheij, F. and Absil, F.G.J. (2006) Mission Oriented Sensor Management.: *Proc. Cognitive systems with Interactive Sensors,* Paris.

Boyd, J.R. (1987-1992) A Discourse on Winning and Losing. Unpublished briefing notes. Various editions. Available from: http://www.d-n-i.net/richards/boyds_ooda_loop.ppt

De Jong, J.L., Burghouts, G.J., Hiemstra, H., Te Marvelde, A., Van Norden, W.L., Schutte, K. and Spaans, M. (2008) Hold Your Fire!: Preventing Fratricide in the Dismounted Soldier Domain, *Proc. 13th Int. Command and Control Research and Tech. Symp.* Bellevue, WA.

Huizing, A.G. and Bloemen, A.A.F. (1996) An Efficient Scheduling Algorithm for a Multi Function Radar, *Proc. IEEE int. symp. on Phased Array Systems and Technology*, Boston, MA, pp 359-364.

Klein, G.A. and Crandall, B.W. (1996) *Recognition-Primed Decision Strategies*, ARI Research Note 96-36, U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.

Komorniczak, W. Kuezerski, T. and Pietrasinski, J.F. (2000) The Priority Assignment for Detected Targets in Multi-Function Radar, *Proc. 13th Int. Conf. On Microwaves, Radar and Wireless Communications*, Mikon, pp 244-247

McIntyre, G.A. and Hintz, K.J. (1999-I) A Comprehensive Approach to Sensor Management, Part I: A Survey of Modern Sensor Management Systems, *IEEE Transactions on SMC.*

McIntyre, G.A. and Hintz, K.J. (1999-II) A Comprehensive Approach to Sensor Management, Part II: A new hierarchical model, *IEEE Transactions on SMC.*

McIntyre, G.A. and Hintz, K.J. (1999-III) A Comprehensive Approach to Sensor Management, Part III: Goal Lattices, *IEEE Transactions on SMC.*

NATO. AJP - 01(B) (NATO/PfP unclassified): Allied Joint Doctrine.

Strömberg, D., Andersson, M. and Lantz, F. (2002) On Platform-Based Sensor Management, *Proc. 5th Int. Conf. on Information Fusion*, Annapolis, ML, pp 1374-1380.

Van Delft, J.H. and Schuffel, H. (1995) *Human factors onderzoek voor toekomstige commando centrales KM*, TNO-TM 1995 A-19 (in Dutch).

Van Norden, W.L., De Jong, J.L., Bolderheij, F. and Rothkrantz, L.J.M. (2005) Intelligent Task Scheduling in Sensor Networks, *Proc. 8th Int. Conf. on Information Fusion*, Philadelphia.

Modelling Human-like Visual Perception for Intelligent Multi-modal Information Fusion

Coen Stevens, Theo Hupkens & Léon Rothkrantz

Introduction

Military sensors are being used to create situational awareness. In a process of 'multi-sensor data fusion', a sensor grid containing a multitude of similar and different sensor types contributes to the overall awareness by gathering and combining input data. Such data fusion has been applied in numerous military applications including ocean surveillance, air-to-air defence, battlefield intelligence, surveillance and target acquisition, and strategic warning and defence [Hall, 2001]. Regarding the sensor grid there are several recent developments, i.e., sensor types become multimodal and mobile sensor deployment will be increasingly autonomous. Multimodal means using multiple modalities, which are different types of physical phenomenon that can be sensed, such as light and sound. In terms of military applications one can think of a combination of radar and electro-optical systems, or electro-optical combined with acoustic. Non-military examples of incorporating multimodal fusion include: enhancing automatic speech recognition with visual features, and person identity verification.

The aim of our research is to design and implement an autonomous and adaptive surveillance system based on a video and acoustic sensor. In order to achieve our goal, we need to implement a suitable data fusion framework and fitting fusion techniques. The fusion of the two modalities: vision and audio, has to solve the problems of ambiguity, redundancy and synchronicity in a seamless manner. The idea is that the surveillance system takes over the task of the human observer, which means interpreting the scene and spotting for aggressive or other illegal activities that will have to be reported back to surveillance personnel who can then take the appropriate action.

In order to achieve autonomous surveillance with a multimodal, intelligent sensor we believe that understanding and modelling human perception is at the crux of making intelligent context sensitive systems that try to make sense out of an overwhelming amount of data coming in through their sensors. In this article we will focus purely on the visual part of our fusion model and present our computational model for visual perception including the results we have so far.

Modelling human perception

Humans unconsciously utilize audiovisual information fusion continuously. For example, when listening to a speaker, we also tend to look at his or her lip movements (and other non-verbal signs, like gestures), which help us to improve speech recognition by utilizing the complementary information in vision and audition. Not only do we receive more information using multiple senses, but multimodal processing can help us to resolve ambiguous information within any single modality. This enhances our situational understanding and awareness.

Neurological evidence suggests that multimodal fusion is only done at a higher level following the perception of each of the separate modalities. However it appears that high-level perception, the level at which concepts and representations come into play, is not separable from low-level perception (basic processing of incoming data), being deeply intertwined [Chalmers et al., 1992]. Not only will low-level perception influence high-level concepts (bottom-up), also the conceptual influence keeps perception flexible given any context (top-down). For example when we have prior knowledge of a situation, say we are given a picture and are told in advance that there will be a man in the picture, we use this high-level concept of 'a man' to group the low-level input by looking for 'man-features'. Or another example: when we appear to see a face, we tend to interpret the features in the face as eyes, mouth and nose, even when the picture is so unclear that we would not have recognized a nose if the area of the nose would have been presented in isolation. This is the power of context, which is the interpretation of lower level features given the higher-level concepts (e.g., face). Now this is not something exclusively for visual or auditory perception, it also works between these two modalities. When you hear meowing you expect to see a cat! Most work to date in visual and auditory perception has been targeted at either bottom-up or top-down processing. The main challenge for future models of perception is the integration of such top-down influences with bottom-up processing [Riesenhuber and Poggio, 2000]. We believe that such an integration can be accomplished by modelling auditory, visual and audiovisual perception on an underlying theory of 'self-organization' and 'emergent behaviour', which will be explained in detail in the following sections.

Background

Emergent perception

Emergent behaviour can be found in nature among many species where their local actions and interactions result in a global behaviour of the entire group which is novel with respect to the behaviour of every single member of the group. For example ants leaving pheromone trails while gathering food, lead to an effective path-finding strategy of food sources for the entire group of ants that follow the strongest (reinforced) trails.

A lot of psychophysical and neurological evidence suggests that perception deploys emergent mechanisms resembling the above mentioned emergent behaviour. The emergent properties of perception were shown to exist by the Gestalt psychologists and their Gestalt theory, which started in 1921 with the Max Wertheimer founding paper [Wertheimer 1923]. Gestalt theory was a reaction to the established notion of structuralism introduced by W. Wundt in 1879. Structuralism stated that perception is built up from atomic elements, called sensations, which together by mere addition of all elements constitutes the overall perception. Gestalt theorists on the other hand believed that the whole can be different than the sum of its parts. They emphasized the interaction of the parts and the organizational process as a dynamic process. Gestalt theorists often describe perception as a self-organizing system that spontaneously takes on the 'best' or simplest arrangement in given conditions. During the process of self-organizing perception Gestalts (organized wholes) emerge from the data gathered by our sense organs. Gestalt psychologists provide a theoretical framework based on psychophysical experiments with perceptual laws of organization with which they emphasize the interaction of the parts and the organizational process as a dynamic process.

The following are some of the Gestalt laws of organization (see Fig. 1 for an illustration of Gestalt principles):

- Proximity: when objects are close to one another we tend to perceive these as a group.
- Similarity: when objects look similar we tend to perceive one object rather than separate objects.
- Good continuation: when the eye tends to be led from one (series of) objects to another one, we tend to perceive these as a group.
- Closure: of several possible perceptual organizations, ones yielding "closed" figures are more likely than those yielding "open" ones.
- Orientation and symmetry: objects oriented with horizontal and vertical axes, or ones that are symmetric, are more often perceived as figures.

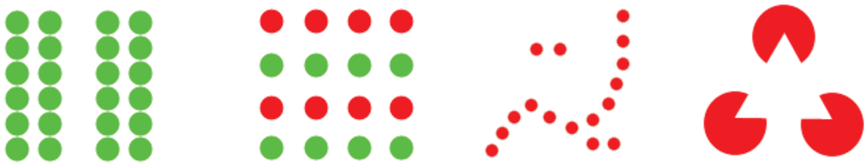

Figure 1. Gestalt principles from left to right, proximity (two groups), similarity (four groups), good continuation (three groups, where one group forms a continuous curved line) and closure (forms a closed triangle)

Camouflage is obviously related to colour and patterns, but also to Gestalts. For example, in Fig. 2 it is according to the law of similarity that the same colour without a border merges into its environment, or as the Gestaltist M. Wertheimer puts it: "If an object is to be hidden by blurring its boundaries, then it is important that besides the colouring, its texture and fine detail are matched to those of its environment."

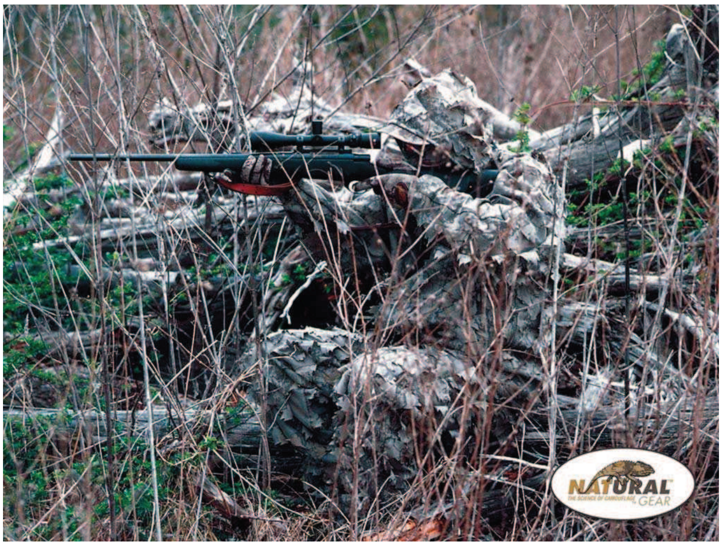

Figure 2. Camouflaged soldier (from natural gear: http://www.naturalgear.com/backgrounds/natgear/1024/3a.jpg)

In Fig. 3 we can see another example of camouflage. Here the ship is painted with random areas of black and white.

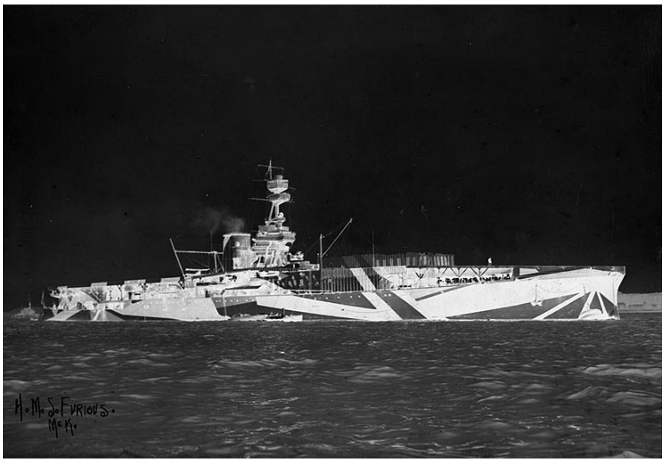

Figure 3. HMS Furious, (British Aircraft Carrier, 1917-1948)
Source: http://www.bobolinkbooks.com/Camoupedia/DazzleShips.html

The texture pattern breaks up the visual outline of the ship when it is seen across the water, and makes it more difficult to tell which way the ship is heading, and to discriminate the different parts of the ship. This type of 'camouflage' does not hide the object from the viewer, but dazzles him.

To summarize, perception as in organizing the input into coherent subsets containing a single object or structure, is the result of competition and cooperation of laws of organization, rather than mathematical bottom-up segmentation. We like to see the organization laws as grouping pressures which try to push and group the input into a particular arrangement. Interaction between grouping pressures at the lowest level of perception give rise to emergent coherent structures (objects), which are novel with respect to the individual cues (e.g., pixels). So it is legitimate to say that the whole is more than the sum of its parts. The reason why interaction among grouping pressures is such a key ingredient arises from the necessity for dealing with the contradictory and incomplete set of cues present at any real-world input caused, among other things, by occlusions, distortions, and reflections. By letting these pressures actively push each other with no centralized interference, structures may emerge, which amount to a reconstruction of the shared fate of the constituent elements.

Computational models of Gestalt principles

Researchers have designed computational models for several Gestalt principles separately, e.g., good continuation, closure and organized contours [Desolneux et al., 2003]. However many have taken a strictly mathematical bottom-up approach, and computed absolute thresholds of meaningful groupings, where they neglected the crucial top-down (contextual) pressures and dynamic interaction among grouping pressures. Grouping pressures (like the Gestalt laws) on their own do not create strong coherent structures. Instead only those supported by an abundance of evidence by other grouping pressures constitute coherent groupings. Take for example the left dot-pattern in Fig. 4, which contains a dotted line that can be easily recognized by humans. In the presence of lots of non-aligned dots it becomes much more difficult to observe the same line. Here seeing the alignment would not be the result of a single 'line-detection' grouping pressure, but is the result of a myriad of grouping pressures that interact and exploit redundant information. Different grouping pressures that propose similar groupings, provide more (redundant) evidence for a coherent strong structure. Examples of such grouping pressures are proximity, good-continuation, regular-orientation and regular-distance. Alternatively (bottom-up) mathematical line detection algorithms could quite easily find the same line in the left dot-pattern of Fig. 4. However they would also still find the same line when more random points are added to the same example, even when humans would no longer see the alignment (see the right example in Fig. 4).

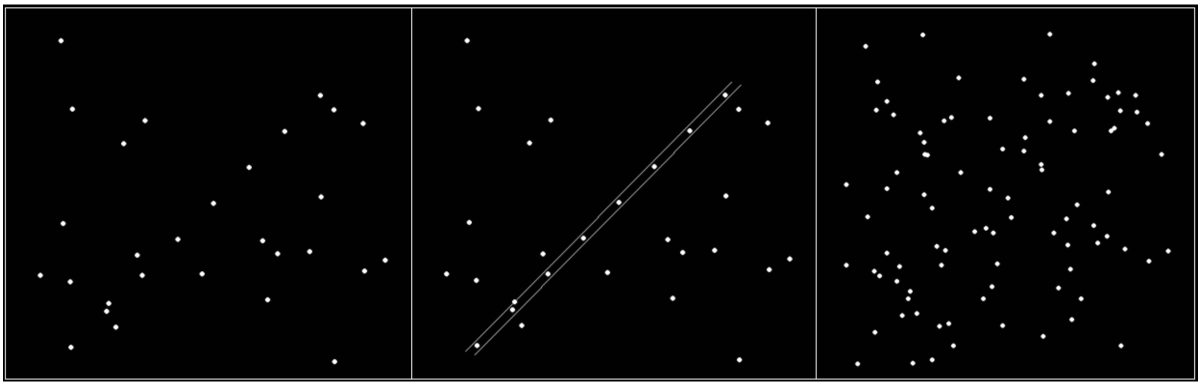

Figure 4. Left: 21 uniformly randomly distributed and eight aligned dots. Middle: this meaningful alignment is detected as a large deviation from the random pattern. Right: same alignment, but with 81 random dots. The alignment is no more meaningful (and it is not seen by the average human observer). In order to be meaningful, it would need at least 11 aligned points. (Examples from Desolneux et al. 2003).

We propose a computational model for visual perception based on the general underlying theory of implementing perception by self-organization [Stevens et al., 2008], which is founded on "The Ear's Mind", an architecture that supports emergent processes, self-organization, and context sensitivity, for the primitive perception of sound [Dor, 2005]. By implementing several visual grouping pressures that utilize the emergent architecture, we will demonstrate their importance and necessity for a computational model of visual perception. Our preliminary results agree with expected human visual grouping behaviour and support our ongoing work on audiovisual fusion.

The Ear's Mind

The Ear's Mind theory, offers a general architecture for simulating emergent sensory perception and specifically for the segregation of auditory scenes [Dor, 2005]. The model of The Ear's Mind was inspired by the 'Copycat' model [Mitchell, 1993; Hofstadter and FARG, 1995]. The Copycat computer program [Mitchell, 1993] models the mechanisms of analogy-making in a letter-string micro-domain. It was designed to be able to discover insightful analogies, and to do so in a psychologically realistic way. In the Ear's Mind, Copycat is abstracted from its original micro-domain and specific sort of analogy-making paradigm. The Ear's Mind is designed to model the unconscious, automatic auditory grouping pressures in humans. Such pressures, it seems, steer the perception of sound by cooperative and competitive interactions, resulting in the grouping of sound elements into context-sensible entities. These are the pressures we talked about in the previous sections. A software prototype, simulating the most basic functionality of the proposed model has already been implemented and presented with sound excerpts of standard psychoacoustic experiments [Bregman, 1990]. Preliminary results agree with expected human auditory grouping behaviour [Dor and Rothkrantz, 2008].

A computational model for visual perception

Based on the same architecture as the Ear's Mind, our emergent system works as a non-supervised collection of independent local primitive agents which represent and act as local grouping pressures (e.g., proximity and regularity) that will try to force a specific grouping onto the input. These agents compete and cooperate to build or destroy bridges in the data-landscape they work on, resulting in the creation of high-level structures out of low-level input. Different grouping pressures that propose similar groupings, provide more evidence for a coherent strong structure. We take the visual Gestalt laws of organization as our initial starting point for modelling several different grouping pressures, but we do not take only the Gestalt laws as an exhaustive list of possible pressures. Before delving into more detail on the grouping pressures we first turn to the overall architecture of our visual perception model.

Architecture

The architecture, illustrated in Fig. 5, is based on four major building blocks: the Pre-processor, Workspace, Slipnet, and Coderack containing Agents.

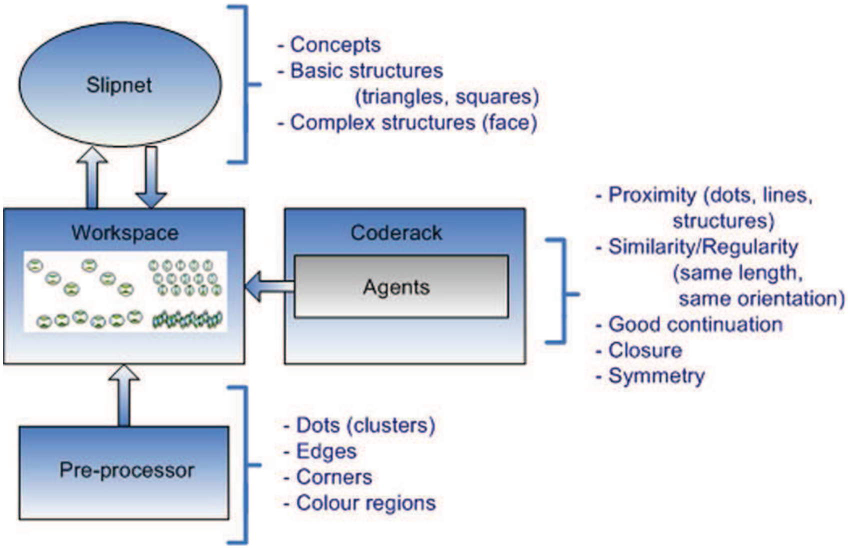

Figure 5. The architecture of our visual perception model

The *Pre-processor* analyses the input image and produces a list of salient cues, containing cue type, location and any other properties needed for a cue's definition. It is up to the pre-processor to fill the workspace with the most primitive cues and not with higher level interpretations or groupings of primitive cues. We do not propose an exhaustive list of primitive cues, but rather keep the option open to include more different cues as we go along. Currently we have only a single type of cue, namely dots, to study and model organizations of dot patterns. Later on, for more complex input images we certainly need to resort to other primitive cues, e.g., density, gradients, colour and edges. One thing we have to keep in mind is that these artificially constructed dot patterns are in a way more

difficult than real-life images, in the sense that in complex images we can find a lot more redundant information for building coherent structures.
The *workspace* is where the actual construction is taking place with the building of perceptual structures on top of the cues. When the workspace is filled with cues, we are ready to start launching local agents. We have implemented the following four different types of agents, which are described in more detail in the following sections: Proximity, Regular-orientation, Good-continuation and Regular-distance. Over time, through the actions of these agents, cues in the workspace gradually acquire various descriptions, and are linked (bonded) together by various perceptual structures. It is important to see that the strength of the architecture lies in the combination of multiple agents and their interaction, and not so much in any single agent. One imperfect agent that suggests a particular grouping that is in conflict with the grouping of a structure built by many other agents supporting each other, will by no means affect the overall good outcome of the system. Hence we do not try to build perfect agents, but we want to find and gather as much grouping evidence as possible. Initially we randomly launch the agents on the workspace, which means that the system works strictly in a bottom-up fashion. Later on the architecture provides in the necessary top-down influences to direct the launching of agents in an appropriate way and to focus on the most relevant cues and structures given the context of the input image. The agents will be placed in the so-called '*Coderack'*, which is a waiting room filled with agents that will investigate possible structuring in the workspace and making probabilistic decisions. Agents are stochastically selected from the Coderack (the name 'Coderack' is taken from Copycat, where agents were called 'Codelets'). For example if we would not incorporate top-down pressures and in a particular case start and continue with a high portion of the agents being regular distance agents, one is bound to find regularity in the end, even though we might not perceive this regularity due to stronger structures in the context, which are not found because we only focused on finding regularity in the first place. Therefore we need to regulate the agents to be launched. If like in the previous case regularity is hard to find, then less agents need to be launched to search for this type of grouping, especially when another structure based on non-regular evidence is being formed.

The *Slipnet* is responsible for the top-down influences, which is a network of interrelated concepts, where each concept is represented by a node and is surrounded by potential associations and slippages (changing from one concept into another). Conceptual relationships represented as links have a numerical length, which resembles the 'conceptual distance'. Conceptual links in the Slipnet adjust there lengths dynamically as the conceptual distances gradually change under the influence of the evolving structure in the workspace. In the Slipnet each of the concepts can become active when instances of them are noticed in the workspace. Also agents can provide feedback to the workspace by creating a top-down pressure to look for further instances of active concepts. Furthermore, concepts can spread activation to their neighbours.

Building bonds and relations

On the workspace we distinguish two types of bonds: 'cue-bonds' and 'relation-bonds'. The cue-bond is proposed between two cues, like in the middle example of Fig. 6, where we have the three dot cues, from the left example of Fig. 6, and a bond represented by an arrow that starts in the most right dot-cue and points to the middle one. What the bond represents is a local view from the right cue stating that it groups together with the middle cue, based on the grouping pressure that proposed and built the bond. In our model 'Proximity' is an example of a grouping pressure that constructs such cue-bonds. Some other grouping pressures, like regular distances, are not cue-bonds, but bonds among the distance relation between two cues and the distance relation between two other cues. If they have equal distances, then we can speak of regular distances. To express these groupings between two sets of cues we use the relation-bond. For example in the right dots-pattern of Fig. 6 we displayed a bond between two relations, where the two dashed lines between the first and second dot, and the second and third dot represent two relations, which are bonded by the grey pointing arrow. Distance and orientation are the two relations we used in our model. For a relation-bond to be built we first need to build the relations on the workspace.

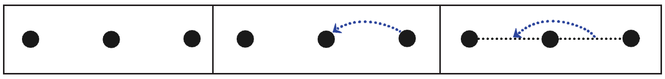

Figure 6. Left: three dot cues. Middle: Cue to Cue bond.
Right: Relation to Relation bond.

The actual proposing and building of bonds is split into work for two different types of agents: *scouts* and *builders*. Scouts search the workspace for cues to bond, and the builder-agent builds the bond. Initially we launch 'Propose Bond'-scouts, which follow two rules:

1. Land on two or three cues.
2. IF fitting pressure description THEN Propose bond and put Bond Builder in the coderack ELSE terminate.

Next, when the Bond Builder is launched it acts as follows:

1. Check for resistance to bond. It is possible that existing bonds oppose building the proposed bond.
2. IF NO resistance THEN build bond ELSE fight: Resulting in either building or deleting the proposed bond.
3. IF proposed bond is built THEN post Bond Extender scout.

The Bond Extender scout on its turn does the following:

1. Lands on the bond to be extended.
2. Checks for extensions to propose new bonds.
3. IF proposing a new bond THEN put Bond Builder in the coderack ELSE terminate.

Now that we have explained the general bond building mechanism, we can move on to the specific grouping pressures.

Grouping pressures
We have implemented the following four grouping pressures, which are modelled after the Gestalt laws of visual perception: *Proximity, Regular-orientation, Good-continuation and Regular-distance*. It is important to remember that the strength of our model lies in the

combination of multiple grouping pressures and their mutual supporting evidence, and not so much in any of them in isolation. Different grouping pressures that propose similar groupings, provide more (redundant) evidence for a coherent strong structure. Next we will describe each implemented grouping pressure in more detail including their grouping results on our alignment example from Fig. 4.

Proximity

The Proximity scout proposes and builds bonds between cues based on the distance between cues. The purpose of the Proximity agent is to bond each cue from a local perspective to other cues which are the closest. When we land with our proximity scout on a dot (cue) we take this cue as the centre point of a circular search zone for which we make a list of all the cues within this zone. For each cue we find in our search zone we calculate the Euclidean distance to the centre cue, and use these distances to set up a probability for being a candidate for a proximity bond. The shorter the Euclidean distance the higher the chance the cue will be chosen to build a proximity bond. Strong proximity relationships are between those cues that both have proximity bonds that point to each other (two-way proximity), which shows that from the local perspectives of each of the two cues the other cue is proximate.

The results on alignment examples are displayed in Fig. 7 after 100 scouts were launched onto the workspace. In the left result we can see many two-way proximity bonds including bonds between the 8 dots (see Fig. 9) that form the visible line among 21 uniformly randomly distributed dots. Only the two bottom-left dots are not bonded together due to another (random) dot that lies really close to the alignment. This suggest that there is proximity evidence for grouping some parts of the line together, but solely on proximity one would not perceive the line. It is interesting to see (although quite messy) that based on proximity the alignment in the right result of Fig. 7 has no support whatsoever and is totally disturbed by interfering close by random dots, which is just as one would expect to see based on proximity.

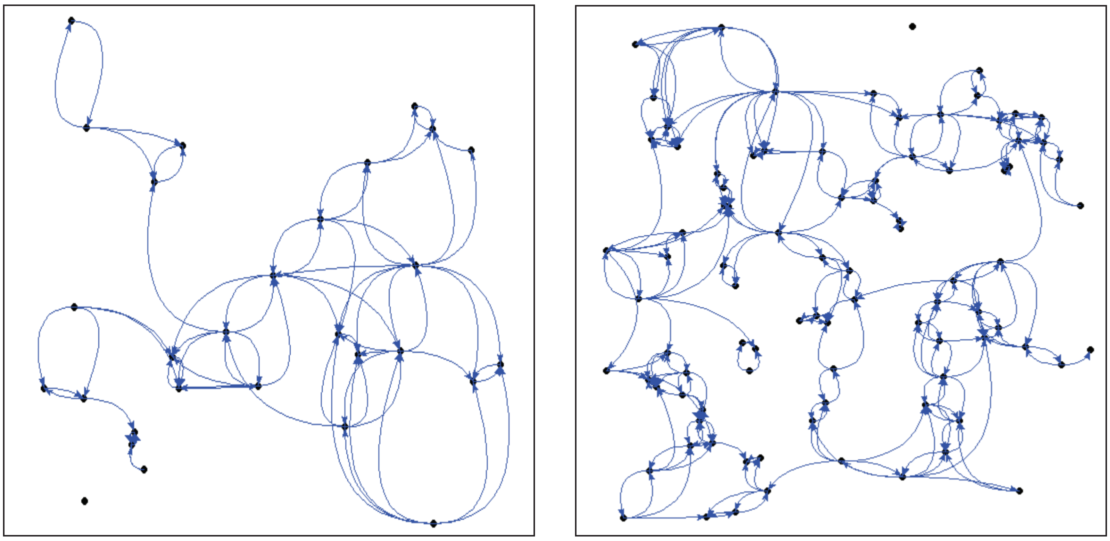

Figure 7. Left: Proximity bonds on 21 uniformly randomly distributed and eight aligned dots. Right: Proximity bonds on 81 uniformly randomly distributed and the same eight aligned dots.

Regular-orientation

The Regular-orientation scout proposes and builds bonds between orientation-relations that have the same orientation and share one cue, which essentially means a straight line through three dots. We explain how the Regular-orientation scout operates by the use of the example given in Fig. 8, where we initially start with three dots (leftmost dot-pattern).

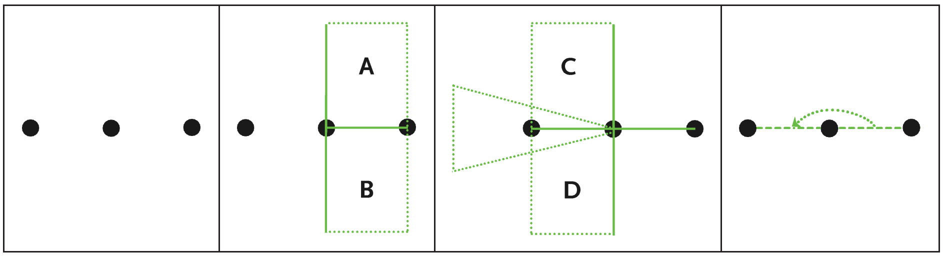

Figure 8. Leftmost: three initial dots. Middle left: free zone check. Middle right: free zone check and triangle search zone. Rightmost: Regular-orientation bond.

First the agent lands on a random dot cue, which is in this example the dot (cue) to the far right and we take this cue as the centre point of a circular search zone for which we make a list of all the cues within this zone. The diameter of the search-zone should be sufficiently large, not to exclude the finding of lines over large distances. We find two cues and for both cues we calculate the distance to the rightmost cue, and use these distances to set up a probability for being a candidate for the second dot on the line. The shorter the distance the higher the chance the cue will be chosen (just like we did with the proximity scout). Say we choose the middle point as the second dot. Now we check for free zones that need to be free of interfering dots, illustrated in the middle left example by two rectangular zones (A and B). The zone's width is proportional to its length, which is the distance between the first and second dot. If both A and B would contain any dots, then the scout will terminate. On the other hand if only one of them includes a dot or if they are both free of them, then we continue the search for a third dot. The reason why we introduce the concept of free zones, is that it helps to home in on 'clear' lines by avoiding dense clusters of dots. In search for the third dot the scout constructs a triangle search zone in the direction from the first to the second dot, starting from the second dot (see middle right example). The length and width of the triangle are proportional to the distance between the first and second dot. From all the dots found in the triangle zone we calculate the distances to the middle cue and use these distances to set up a probability for being a candidate for the final third dot on the line. In our example we find only one dot, and also here we check for interfering dots between the second and third dot, just like we did between the first and second dot with rectangular zones (C and D). We have three conditions under which we abort proposing a regular orientation bond, because under these conditions both sides of the alignment would have interfering dots:

- If there are cues in rectangle A and D.
- If there are cues in rectangle B and C.
- If there are cues in rectangle C and D.

If none of these conditions apply then the scout proposes to bond the orientation relation between the first and second cue, and the same relation between the second and third cue as shown in the rightmost example of Fig. 6.

The results of the regular orientation scout on alignment examples are displayed in Fig. 9 after the actions of 100 scouts on the workspace. In the left result we can see that the scout for this grouping pressure easily finds and groups the alignment of dots together. Additionally it finds even more dots that form other alignments. These alignments are correct. However, when we look at the total dot pattern, we are not drawn to these other alignments and will not mark them as something interesting. This is just one opinion of one type of agent, which unless it is supported by any other grouping pressure remains a weak structure. In the right result of Fig. 9 we see that the found alignments are all over the place and none seem to fit the 'hidden' 8-dot alignment, which matches expected human visual grouping in this particular example.

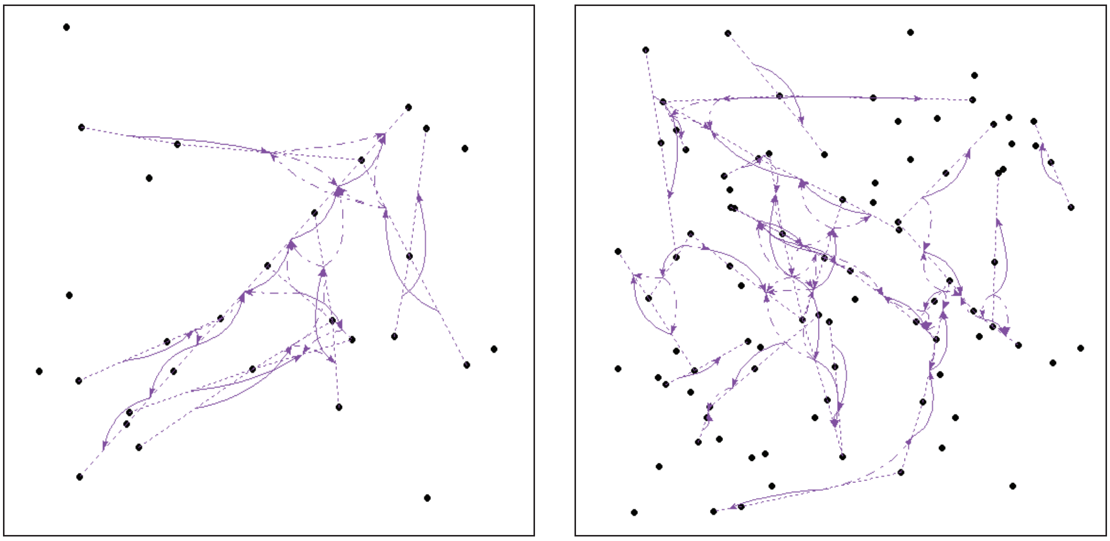

Figure 9. Left: Regular-orientation bonds on 21 uniformly randomly distributed and eight aligned dots. Right: Regular-orientation bonds on 81 uniformly randomly distributed and the same eight aligned dots.

Good continuation

The Good-continuation scout works in an almost identical way as the Regular-orientation scout, working also with orientation relations. Only where the Regular-orientation scout spots straight lines, the Good-continuation scout finds the best continuation of a line, which could be slightly curved. For this behaviour the scout follows the same steps as the Regular-orientation scout, only allows the triangle search zone for the third dot to be much wider and has a different selection criterion for the best candidate dot in the triangle zone. The selection criteria is no longer based on being nearer to the second dot (the point where the triangle cone begins), but based on best fitting of the orientation between the first and second dot. The results of the good-continuation scout are presented in Fig. 10 and resemble the results of the regular-orientation scout, with only minor differences.

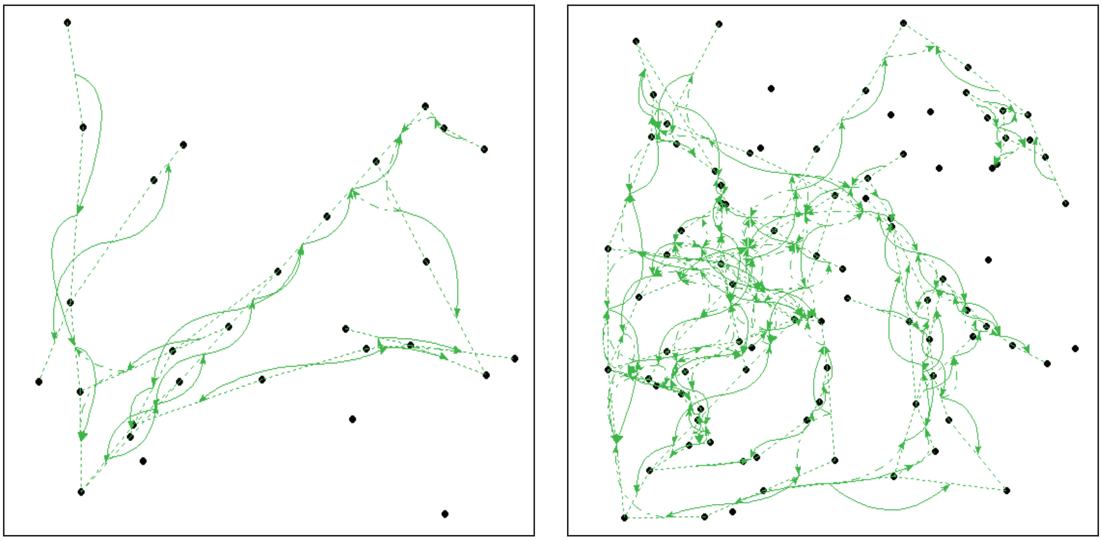

Figure 10. Left: Good-continuation bonds on 21 uniformly randomly distributed and eight aligned dots. Right: Good-continuation bonds on 81 uniformly randomly distributed and the same eight aligned dots.

Regular distance

Finally the fourth scout we have implemented, the Regular-distance scout tries to bond cues together that have the same inter distance. With the help of the example in Fig. 11 we will demonstrate how this agent operates.

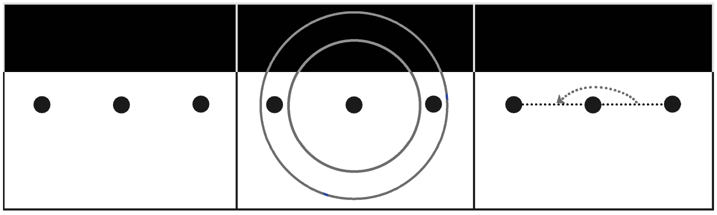

Figure 11. Left: three dot cues. Middle: margin space. Right: Regular-distance bond.

First we land on a random cue in the workspace, which in this example is filled with three dot cues (leftmost dot-pattern). The agent lands on the middle cue and we use this dot as the centre point of a circular search zone for which we make a list of all the cues within this zone, and find two other cues (the leftmost and the rightmost).

For both found cues we calculate the distance to the middle cue, and use these distances to set up a probability for being selected for the second step. The shorter the distance the higher the chance the cue will be chosen. Say we would have chosen the far right cue to perform the second step of the agent, which is finding other cues that have the same distance to the middle cue. We make a list of all the cues with the same distance, given a small error margin, illustrated in the middle figure by two circles. In our example we find the leftmost dot within the margins. The next step the scout proposes to bond the distance relation between the middle and rightmost cue, and the same relation between the middle and leftmost cue as shown in the rightmost example of Fig. 11.

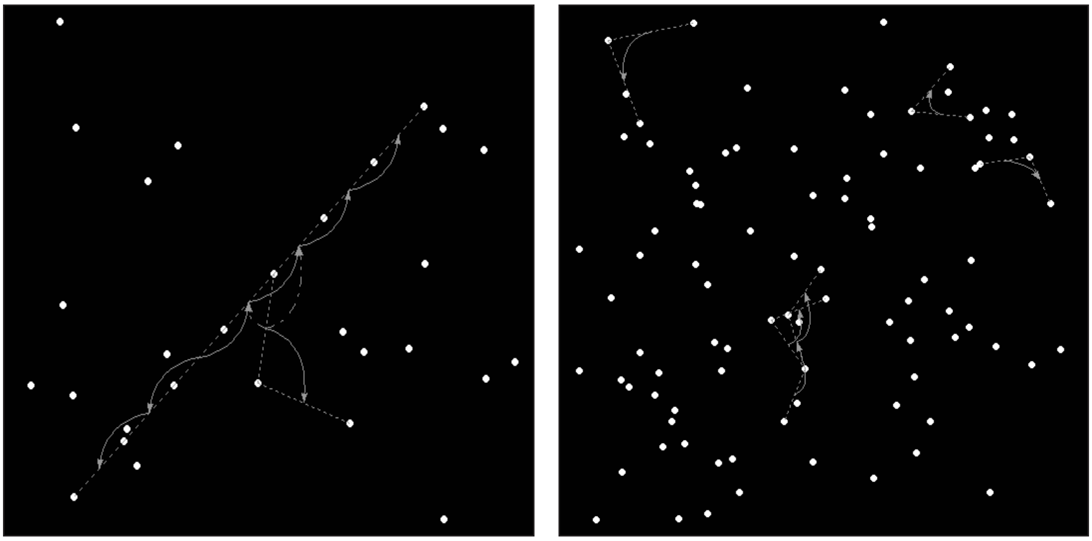

Figure 12. Left: Regular-distance bonds on 21 uniformly randomly distributed and eight aligned dots. Right: Regular-distance bonds on 81 uniformly randomly distributed and the same eight aligned dots.

The results of the Regular-distance scout are presented in Fig. 12, and just like with Regular-orientation and Good-continuation, this scout flawlessly discovers the alignment and this time it is almost the only thing it finds apart from one other bond. Furthermore, as expected the scout finds none of the 'hidden' 8-dots in the Bottom example.

Joint Grouping pressures

Fig. 13 depicts the combined results of the Regular-orientation, Good-continuation and Regular-distance grouping pressures. The proximity grouping pressure was left out for clarity. The alignment of mutually supportive grouping pressures can be clearly seen in the left result of Fig. 13. From this example, the advantage of using mutually supportive grouping pressures is contrasted with the interpretation power of each grouping pressure on its own. Consequently, only those bonds supported by multiple grouping pressures may lead to the formation of higher level structures.

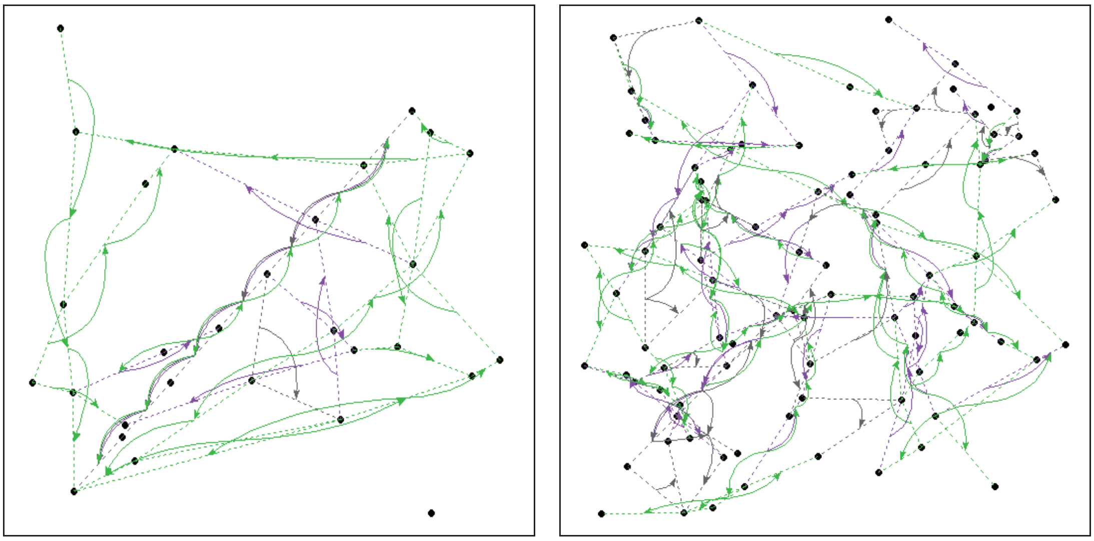

Figure 13. Combined results where purple-bonds = regular orientation, green bonds = good continuation, grey-bonds = regular distance. Left: Combined results on 21 uniformly randomly distributed and eight aligned dots. Right: Combined results on 81 uniformly randomly distributed and the same 8 aligned dots.

Future work

A plan for a second implementation phase has been devised for taking the computer program closer to the proposed theoretical model both by extending the cue and agent repertoire and by implementing higher-level capabilities. In addition, the visual perception model together with the Ear's Mind is used as a template for implementing an audiovisual fusion model for multimodal perception (see [Stevens et al., 2007]). Such an audiovisual model is expected to enhance the capabilities of real-world scene segmentation in comparison with single modality models. A working model for audiovisual perception can later be augmented and combined with other input data, which is foreign to human perception, like for instance infrared and echolocation. Our model will finally be used to construct an autonomous and adaptive surveillance system based on a video and acoustic sensor. Applications of such a system are harbour protection and battlefield surveillance (detection and recognition of friend or foe).

Conclusions

The aim of our research is to design and implement an autonomous and adaptive (context sensitive) surveillance system based on a video and acoustic sensor. In our approach we try to implement a working model of human audiovisual perception, because we believe that understanding and modelling human perception is at the crux of making intelligent context sensitive systems that try to make sense out of an overwhelming amount of data (e.g., smart surveillance systems). Our initial goal, which we described in full detail, was to focus on visual perception and to implement a working model of primitive visual perception, integrating top-down influences with bottom-up processing, and mimicking perceptual grouping behaviour of human subjects. We proposed the visual perception model as a novel emergent, self-organizing model, supported by neurological and psychological evidence. The model consists of an open architecture allowing the addition of new features, pressures and interaction methods, making it possible to define more agents and extend the model's capabilities. Following the design phase, the model was implemented as a software prototype, and was used for testing Proximity, Good-continuation, Regular-orientation and Regular-distance grouping pressures. Results so far show that the implemented model forms a promising foundation for further research and expansion for dealing with more complex images.

References

Bregman, A.S. (1990). *Auditory Scene Analysis: The Perceptual Organisation of Sound,* 2nd paperback ed. 1999, MIT Press, Cambridge.

Chalmers, D.J., French, R. M., and Hofstadter, D. R. (1992) High-Level Perception, Representation, and Analogy: A Critique of Artificial Intelligence Methodology. *Journal of Experimental and Theoretical Artificial Intelligence* 4, 185–211.

Desolneux, A., Moisan, L. and Morel, J.-M. (2003) Computational Gestalts and Perception Thresholds. *Journal of Physiology-Paris* 97, 311–324.

Dor, R. (2005) *The Ear's Mind: A Computer Model of the Fundamental Mechanisms of the Perception of Sound,* Technical report 05-16, Delft University of Technology.