Hierarchical Selectivity for Object-based Visual Attention

Yaoru Sun and Robert Fisher
yaorus,rbf@dai.ed.ac.uk

Division of Informatics, University of Edinburgh, Edinburgh, EH1 2QL, UK

Abstract. This paper presents a novel “hierarchical selectivity” mechanism for object-based visual attention. This mechanism integrates visual salience from bottom-up groupings and the top-down attentional setting. Under its guidance, covert visual attention can shift not only from one grouping to another but also from a grouping to its sub-groupings at a single resolution or multiple varying resolutions. Both object-based and space-based selection is integrated to give a visual attention mechanism that has multiple and hierarchical selectivity.

1 Introduction

Machine vision research has recently had an increased interest in modelling visual attention and a number of computable models of attention have been developed [8, 13, 1]. However, these models are all space-based and do not account for the findings from recent research on object-based attention (see [2, 12] for object-based attention views). These space-based attention models may fail to work in environments that are cluttered or where objects overlap or share some common properties. Three different requirements of attention are immediately identifiable: 1) attention may need to work in discontinuous spatial regions or locations at the same time; 2) attention may need to select an object composed of different visual features but from the same region of space; 3) attention may need to select objects, locations, and/or visual features as well as their groupings for some structured objects. For applying attention mechanisms in real and normal scenes, a computational approach inspired by the alternative theory of object-based attention is necessary. In contrast to the traditional theory of space-based attention, object-based attention suggests that visual attention can directly select discrete objects rather than only and always continuous spatial locations within the visual field [4, 6, 12]. A complete computable model of object-based attention is still an open research area. Moreover, as suggested in [12], “Attention may well be object-based in some contexts, location-based in others, or even both at the same time.” Inspired by this idea, here we present a “hierarchical selectivity” mechanism which is a part of our computable model of object-based attention (not published in this paper). This mechanism guides (covert) attentional movements to deal with multiple selectivity in a complicated scene. The objects of selection can be spatial locations, objects, features, or their combinatorial groupings. Hierarchical selectivity works on the hierarchical structure of groupings competing for attention and navigates attention shifts between coarse groupings and fine groupings at single or multiple resolution scales. Stimulus-driven and top-down biasing are integrated together. Also, Winner-Take-All (WTA) and “inhibition of return” strategies are embedded within the mechanism. In the following section, the

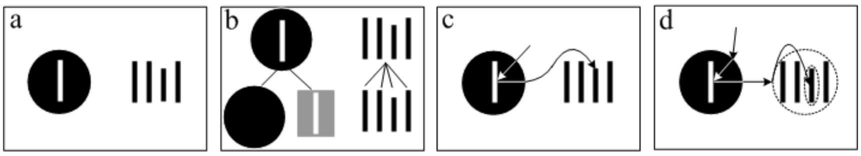

Fig. 1. An example of attention working on hierarchical grouping. a: the original display; b: Two hierarchical groupings obtained from the display c: Possible space-based attention movements; d: Possible object-based attention movements by hierarchical selectivity. Attention firstly selects the grouping consisting of the black circle and white bar and then shifts to the sub-grouping, i.e. white bar. The black bar belonging to another grouping including four black bars is attended after its parent is visited.

background theory used to compute bottom-up salience is briefly introduced. Hierarchical selectivity is presented in Section 3 and experimental results are given in Section 4.

2 Background

The pivotal idea in our solution for object-based attention is the grouping-based salience computation and attention competition (see [14] for detailed implementation). The salience of a grouping measures how different this grouping contrasts with its surroundings and depends on various factors, such as feature properties, perceptual grouping, dissimilarity between the target and its neighbourhood [3, 10]. A grouping is a hierarchical structure of objects and space, which is also the common concept in the literature of perceptual grouping [11, p. 257-266]. A grouping may be a point, an object, a region, or a structured grouping. Figure 1 shows an example of attention working on hierarchical groupings. In this paper, we focus on presenting hierarchical selectivity and assume that the scene has already been segmented into groupings of similar colour, texture, intensity, etc. The input colour image is decomposed into 4 double-opponent colour (red R, green G, blue B, and yellow Y) pyramids, one intensity I pyramid and 4 or 8 orientation pyramids ($u(\theta)$ and $\theta = [0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}]$ or $\theta = [0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2}, \frac{5\pi}{8}, \frac{3\pi}{4}, \frac{7\pi}{8}]$) to create feature maps using overcomplete steerable filters [7, 8]. Then the salience of different groupings at different resolution scales is obtained from these feature maps by the computation of grouping salience. Finally, various groupings compete for visual attention based on the interaction between their bottom-up salience and top-down attentional setting through the hierarchical selectivity mechanism (see Section 3 for details). To save space, the following mathematical description of the grouping salience computation omits the expression of resolution scale. But note that the salience of all groupings is actually calculated at their current resolution and is dynamically varied with different scales and surroundings. Therefore the salience maps (including the salience maps for each grouping) are also multi-scale and dynamical. The discussion of recent psychophysical findings that support the salience computation approach (such as center-surround mechanisms used to encode the salience of visual objects, etc.) are omitted too to save space (see [9, 14] for detailed and extensive discussion). Suppose $\Re$ is a grouping in an image at a given resolution, x, y are any two points in the image, (r, g, b) are the red, green, and blue colour components of the input image. Then the colour chromatic contrast ΔC and intensity contrast ΔI between x and y are calculated as:

$$\Delta C(x,y) = \sqrt{\eta_{RG}^2 RG^2(x,y) + \eta_{BY}^2 BY^2(x,y)}; \quad \Delta I(x,y) = |I_x - I_y|; \quad I_x = \frac{(r_x + g_x + b_x)}{3}; \quad \eta_{BY} = \frac{2\sqrt{B_x^2 + B_y^2 + Y_x^2 + Y_y^2}}{3 \times 255} \quad (1)$$

$$\eta_{RG} = \frac{R_x + R_y + G_x + G_y}{\sum\limits_{x,y} R_{x,y} + \sum\limits_{x,y} G_{x,y} + \sum\limits_{x,y} B_{x,y} + \sum\limits_{x,y} Y_{x,y}}; \quad RG(x,y) = \frac{|(R_x - G_x) - (R_y - G_y)|}{2}; \quad BY(x,y) = \frac{|(B_x - Y_x) - (B_y - Y_y)|}{2} \quad (2)$$

$$R_x = r_x - (g_x + b_x)/2;\ \ G_x = g_x - (r_x + b_x)/2;\ \ B_x = b_x - (r_x + g_x)/2;\ \ Y_x = (r_x + g_x)/2 - |r_x - g_x|/2 - b_x \tag{3}$$

Let $S_{CI}(x,y)$ be the colour-intensity contrast and d_{gauss} be the Gaussian weighted distance between x and y, $\mathcal{NH}_{CI}$ be the neighbourhood surrounding x, $y_i \subset \mathcal{NH}_{CI}$ ($i = 1 \ldots n \times m - 1$, $n \times m$ is the input image size) be any neighbour. Then the colour-intensity salience $S_{CI}(x)$ of x is calculated by:

$$S_{CI}(x,y) = \sqrt{\alpha \Delta C(x,y)^2 + \beta \Delta I(x,y)^2}; d_{gauss}(x,y) = \left(1 - \frac{\|x-y\|}{n-1}\right) e^{-\frac{1}{2\sigma^2}\|x-y\|^2}; S_{CI}(x) = \frac{\sum_{i=1}^{n\times m-1} S_{CI}(x,y_i) \cdot d_{gauss}(x,y_i)}{\sum_{i=1}^{n\times m-1} d_{gauss}(x,y_i)} \tag{4}$$

where $||x-y||$ is the chessboard distance between x and y: $||x-y|| = MAX(|i-h|,|j-k|)$, (i,j), (h,k) are the coordinates of x, y in the current resolution image. The Gaussian scale σ is set to $\hat{n}/\rho$ and $\hat{n}$ is the largest of the width and length of the feature maps at the current resolution. ρ is a positive integer and generally $1/\rho$ may be set to a percentage of $\hat{n}$, such as 2%, 4%, 5%, or 20%, 25%, 50%, etc. α and β are weighting coefficients and we here set them to 1. Define the orientation contrast $C_O(x,y)$ between x and y as:

$$C_O(x,y) = d_{gauss}(x,y)\sin(\overline{\theta}_{x,y}); \qquad \overline{\theta}_{x,y} = \frac{\sum_{j=0}^{\zeta-1} j\varphi \sum_{i=0}^{\zeta-1} u_x(i\varphi) u_y((i\varphi + j\varphi) \mod \pi)}{\sum_{j=0}^{\zeta-1}\sum_{i=0}^{\zeta-1} u_x(i\varphi) u_y((i\varphi + j\varphi) \mod \pi)} \tag{5}$$

where *mod* is the standard modulus operator, ζ is the number of preferred orientations, $\varphi = \pi/\zeta$. When ζ is 4 or 8, φ is $\pi/4$ or $\pi/8$. Let $y_i, (i = 1 \ldots n_k)$ be a neighbour in the distance k neighbourhood $\mathcal{NH}_O(k)$ surrounding x (Distance k neighbourhood has $8k$ neighbours). The orientation contrast $\hat{C}_O(x, \mathcal{NH}_O(k))$ of x to its k-th neighbourhood is:

$$\hat{C}_O(x,\mathcal{NH}_O(k)) = \frac{\overline{C}_O(x,\mathcal{NH}_O(k))}{\xi + \omega_k} \quad \overline{C}_O(x,\mathcal{NH}_O(k)) = \frac{1}{n_k}\sum_{y_i \in \mathcal{NH}_O(k)} C_O(x,y_i) \tag{6}$$

where $\omega_k = n_0 - 1$ and n_0 is the number of different directions within $\mathcal{NH}_O(k)$. ξ is a parameter used to prevent a zero denominator and usually set to 1.

Let m_r be the number of "rings" (one ring consists of the neighbours that have the same distance from their "center" x) in a neighbourhood and $d_{gauss}(k)$ be the Gaussian distance of the k-th neighbourhood to x, w_{ijk} be the value on k-th neighbour "ring" on θ_j orientation map of y_i, n_r be the number of "rings" in the whole neighbourhood of x. Then the orientation salience $S_O(x)$ of x to all of its neighbours is:

$$S_O(x) = \frac{\sum_k \hat{C}_O(x,\mathcal{NH}_O(k)) \cdot d_{gauss}(k)}{(\xi + \omega) \cdot m_r \cdot \sum_k d_{gauss}(k)}; \ \ m_r = \sum_k 1 \text{ and } |\hat{C}_O(x,\mathcal{NH}_O(k))| > 0; \ \ \omega = \sum_j \widehat{H}(\theta_j) \tag{7}$$

$$\widehat{H}(\theta) = \{\widehat{H}(\theta_j)\} = \left\{ \sum_k \frac{|H_k(\theta_j) - \overline{H}(\theta_j)|}{MAX\{H_k(\theta_j), \overline{H}(\theta_j)\}} \right\}; \overline{H}(\theta) = \frac{1}{n_r}\sum_k H_k(\theta); \theta = [\theta_1 \ldots \theta_\zeta]; H_k(\theta) = \sum_{y_i \in \mathcal{NH}_O(k)} w_{ijk}(\theta_j, y_i) \tag{8}$$

Let x_i be an arbitrary component within a grouping $\Re$ (x_i may be either a point or a sub-grouping within $\Re$). Then visual salience S of a grouping $\Re$ is obtained from the following formula:

$$S(\Re) = \gamma_{CI}\sum_i S_{CI}(x_i) + \gamma_O \sum_i S_O(x_i) \tag{9}$$

where γ_{CI}, γ_O are the weighting coefficients for the colour-intensity, and orientation salience contributing to the grouping salience and i indicates all components in the grouping. More detailed mathematical descriptions for the computations of early feature extraction and grouping salience can be found in another paper [14].

3 Hierarchical selectivity

Hierarchical selectivity operates on the interaction between bottom-up grouping salience and the top-down attentional setting. It is concerned with "where" attention is going next, i.e. the localization of the groupings to be attended, not "what" the identification of attended groupings are. Therefore, any top-down control related to recognizing objects or groupings is not considered here. The top-down attentional setting is used as a flag at each "decision point" (control whether to go to the next/finer level of a grouping or not) of each grouping in hierarchical selectivity, which is an intention request of whether to "view details" (i.e. view its sub-groupings at the current resolution scales or finer scales) of a current attended grouping. The competition for attention starts first between the groupings at the coarsest resolution. Temporary inhibition of the attended groupings can be used to implement inhibition of return for prohibiting attention from instantly returning to a previously attended winner. More elaborate implementations may introduce dynamic time control so that some previously-attended groupings can be visited again. But here we are only concerned that each winner is attended once. If continuing to check the current attended grouping, the competition for attention is triggered first among the sub-groupings that exist at the current resolution and then among the sub-groupings that exist at the next finer resolution. Sub-groupings at the finer resolution do not gain attention until their siblings at the coarser resolution are attended. If "no", attention will switch to the next potential winning competitor at the same or coarser scale level. By the force of WTA, the most salient sub-grouping wins visual attention. The priority order for generating the next potential winner is:

1. The most salient unattended grouping that is a sibling of the current attended grouping. The winning grouping has the same parent as the current attended grouping and both lie at the same resolution.
2. The most salient unattended grouping that is a sibling of the parent of the current attended grouping, if the above winner can not be obtained.
3. Backtracking continues if the above is not satisfied.

A more precise algorithmic description of hierarchical selectivity is given in Figure 2.

According to [4], [5], and [6], the competition for visual attention can occur at multiple processing levels from low-level feature detection and representation to high-level object recognition in multiple neural systems. Also, "attention is an emergent property of many neural mechanisms working to resolve competition for visual processing and control of behaviour" [4]. The above studies provide the direct support for the integrated competition for visual attention by binding object-selection, feature-selection and space-selection. The grouping-based saliency computation and hierarchical selectivity process proposed here, therefore, offer a possible mechanism for achieving this purpose.

Two goals can be achieved by taking advantage of hierarchical selectivity. One is that attention shifting from one grouping to another and from groupings/sub-groupings to sub-groupings/groupings can be easily carried out. Another is that the model may simulate the behaviour of humans observing something from far to near and from coarse to fine. Meanwhile, it also easily operates at a single resolution level. Support for this approach to hierarchical selectivity has been found in recent psychophysical research

1. competition begins between the groupings at the coarsest resolution
2. if (no unattended grouping exists at the current resolution) goto step 8;
3. unattended groupings at the current resolution are initialised to compete for attention based on
   their salience and top-down attentional setting;
4. attention is directed to the winner (the most salient grouping) by the WTA rule;
   set "inhibition of return" to the current attended winner;
5. if (the desired goal is reached) goto step 10;
6. if ("view details" flag="no") (i.e. don't view details and shift the current attention)
      { set "inhibition" to all sub-groupings of the current attended winner; }
   if (the current attended winner has unattended brothers at the current resolution)
      { competition starts on these brothers; goto step 2 and replace the grouping(s) by these
        brothers; } else goto step 9;
7. if ("view details" flag="yes") (i.e. continue to view the details of the current attended winner)
      if (the current attended winner has no sub-grouping at the current resolution) goto step 8;
      else { competition starts on the winner's sub-groupings at the current resolution;
             goto step 2 and replace the grouping(s) by the winner's sub-groupings; }
8. if ((a finer resolution exists) and (unattended groupings/sub-groupings exist at that resolution))
      { competition starts on groupings/sub-groupings at the finer resolution; goto step 2; }
9. if (the current resolution is not the coarsest resolution)
      { go back to the parent of the current attended winner and goto step 2; }
10. stop.

Fig. 2. The algorithmic description of hierarchical selectivity

on object-based visual attention. It has been shown that features or parts of a single object or grouping can gain an object-based attention advantage in comparison with those from different objects or groupings. Also, visual attention can occur at different levels of a structured hierarchy of objects at multiple spatial scales. At each level all elements or features coded as properties of the same part or the whole of an object are facilitated in tandem (see [2] and [11, p. 547-549] for further discussion and detailed findings).

4 Experiments and discussion

4.1 Grouping effect and hierarchical selectivity on a synthetic display

Figure 3 shows a display in which the target is the only vertical red bar and no one of the bars has exactly the same colour as another bar. Three bars have the same exact orientation and others are separated by different oriented/colour surrounding bars. (Here we adopt the "orientation" of a bar following psychophysical experiments rather than the known concept in computer vision). If not using any grouping rule, each bar is a single grouping by itself. Then we obtain 36 single groupings. If segmenting the display by the bar's direction, the only structured grouping is formed by the 3 vertical bars (not including any black points in the background) which includes the target (forms one sub-grouping) and other two vertical green bars (forms another two-level sub-grouping). In this way, 34 groupings can be obtained in total: a structured three-level grouping and 33 single groupings formed by other bars respectively. The resulting salience maps of groupings and attention sequences for these two segmentations are given in Figure 3. The background (black pixels), colours, and orientations are all considered in the computation for salience. The top-down attentional setting is set to the free state, so this gives a pure bottom-up attention competition.

The results show different orders of paying attention to the targets. The target grouped with two green bars (see Figure 3 (C1), (C2), (C3), and (C4)) has an advantage in attracting attention much more quickly than the non-grouped target. When competition

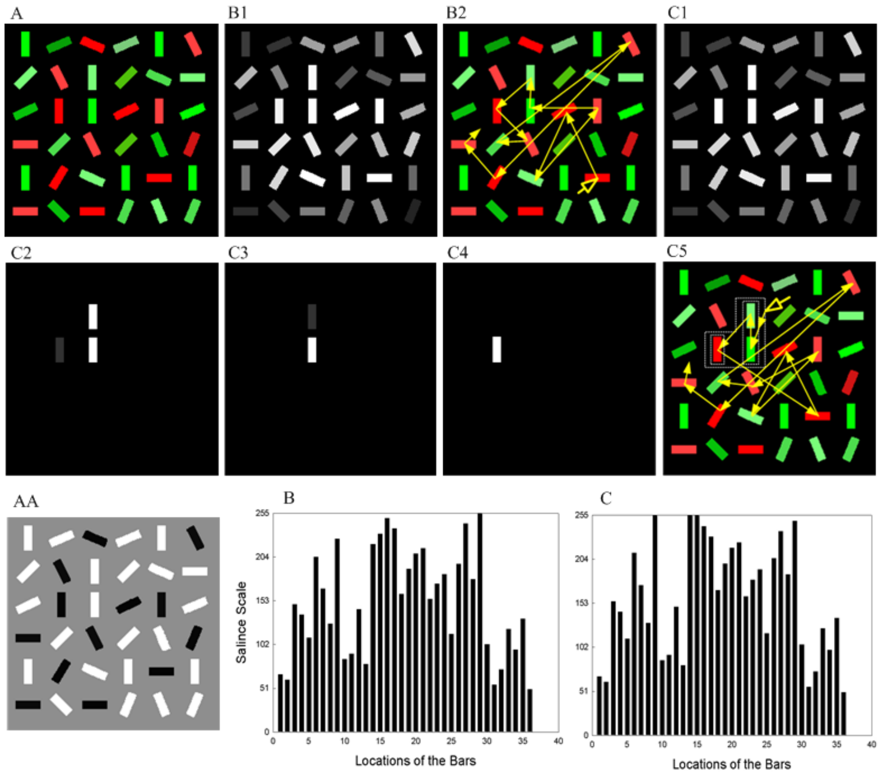

Fig. 3. An example for structured groups and hierarchical selection. In the display the target is the vertical bar at the third row and the second column. A: orignal colour display used in the experiment. AA: monochrome display for A to improve the visibility. All red, green bars are scaled to black, white bars respectively in the grey background. B1: salience map (in shades of grey) in the case of no grouping. B2: attention sequence of most salient bars for B1. C1: salience map in the case of grouping. C2, C3, C4: salience map of the grouped bars. C5: attention sequence of most salient bars for C1. B, C: histograms of B1, C1 respectively. The locations of the bars are simply encoded row column by number 1 to 36, such as the 6 bars in the first columns in B1 and C1 are identified 1 to 6 from left to right. Note the target (bar 9) is attended after 7 movements of attention in B2 but only 3 in C5.

starts, the structured grouping of 3 vertical bars is the most salient and obtains attention firstly. Then the competition occurs within this grouping between the target and another sub-grouping formed by the two vertical but different colour bars. By competition, the target is attended after the two-level sub-grouping is attended. This grouping advantage for attentional competition has been confirmed by psychophysical research on object-based attention [2, 12]. We have applied the model [14] to displays like Figure 3 where we investigated how salience changes with feature (colour, intensity and orientation) contrast, neighbourhood homogeneity and size, target distance, etc. The salience versus changed property curves are similar in shape to the comparable psychophysical results. Thus we claim the model has the desired heterarchical and hierarchical behaviours. More synthetic experiments for testing different behaviours of our model comparing results with those of human observers and other models can be seen elsewhere [14]. However, this research is not intended as a model of human attention, but instead aims at developing a machine vision attention system (inspired by recent psychophysical results) that has the high level of competence observed in humans.

4.2 Performance of hierarchical selectivity in a natural scene

Three colour images shown in Figure 4 are taken using different resolutions from far to near distance (64×64, 128×128, and 512×512) for the same outdoor scene. The scene is segmented (by hand) into 6 top groupings (identified by the black colour numbers: one object grouping 6 and five regions here) and 5 of them are hierarchically structured except grouping 4. In the coarsest image, only grouping 6 (one boat including two people) can be seen. In the finer image, sub-groupings 5-1 and 5-3 within top grouping 5 appear but they lose details at this resolution. The smallest boat (i.e. sub-grouping 5-2 of grouping 5) can only be seen at the finest resolution. The salience maps of groupings during attention competition are also briefly shown in Figure 4 where darker grey shades denote lower salience.

The competition first occurs among the top groupings at the coarsest scene. The most salient grouping 6 therefore gains attention. When giving a “yes” to the top-down attention setting (“view details” flag), attention will shift to the sub-groupings of 6. Two people and the boat then begin to compete for attention. If a “no” is given or after grouping 6 is attended, attention will shift to the next winner grouping 2. If a “yes” is given too to the “view details” flag of 2, attention will first select sub-grouping 2-1 and then shift to sub-grouping 2-2. After attending 2-2, if continuing to view the remainder of 2, attention will shift to the finer resolution to visit 2-3. When grouping 5 is attended, the lake (excluding grouping 6) is visited first and then attention shifts to the finer resolution scene where 5-1 and 5-3 start to compete for attention. In the case of giving a “yes” to the top-down flag of the winner 5-3, attention will shift to the finest resolution scene to check its details. Then attention goes back to the previous finer resolution scene and shifts to 5-1. After that, attention shifts again to the finest resolution scene. Thus the smallest boat 5-2 at the finest resolution is attended. Figure 4 shows the overall behaviour of attentional movements performed on the scene. Using this same scene, when stronger and stronger noise was added above $\sigma = 17$ for Gaussian noise, the order of the attention movements changed. The above results clearly show hierarchical attention selectivity and appropriated believable performance in a complicated natural scene. In addition, although this model is aimed at computer vision applications, the results are very similar to what we might expect for human observers.

Hierarchical selectivity is a novel mechanism designed for shifting attention from one grouping to another or from a parent grouping to its sub-groupings as well as implementing attention focusing from far to near or from coarse to fine. It can work under both multiple (or variable) resolutions and single resolution environments. Here another outdoor scene (figure 5) is used to demonstrate the behaviour of hierarchical selectivity. In the scene, there are two groupings: a simple shack in the hill and a small boat including five people and a red box within this boat in a lake. The people, red box, and the boat itself constitute seven sub-groupings respectively for this structured grouping. The salience maps computed for these groupings are shown in Figure 5 and the sequence of attention deployments is shown in Figure 6. The attention visiting trajectory shown in Figure 6 reveals the reasonable movements of visual attention for this natural scene.

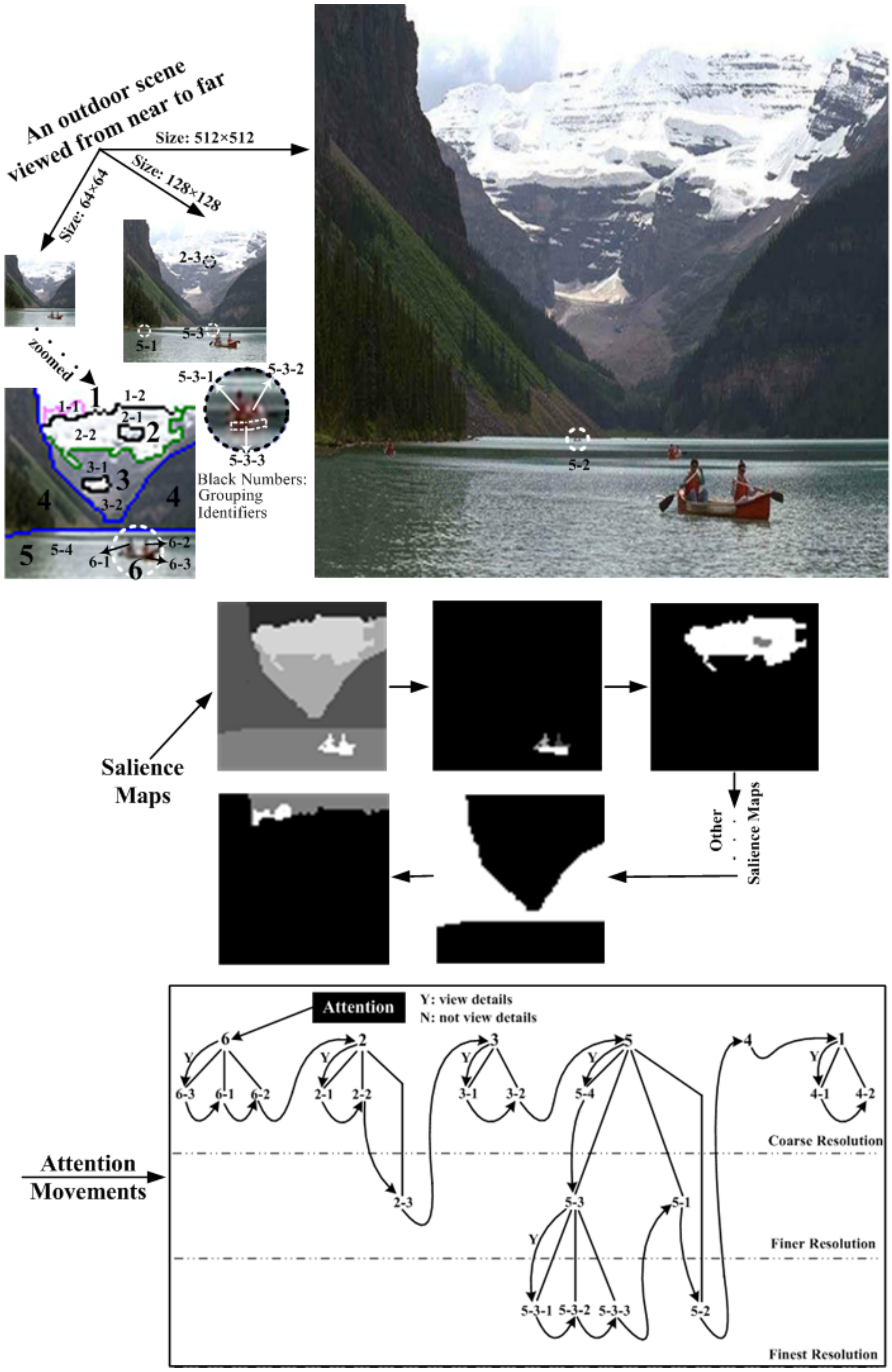

Fig. 4. An outdoor scene taken from different distances. The salience maps and identifiers (black numbers) of different groupings and their sub-groupings are also shown. The dotted circles are used to identify groupings but not their boundaries. The sequence of salience maps used for each selection of the next attended grouping is shown at the left bottom of the figure. Attention movements driven by hierarchical selectivity is shown at the right bottom using a tree-like structure.

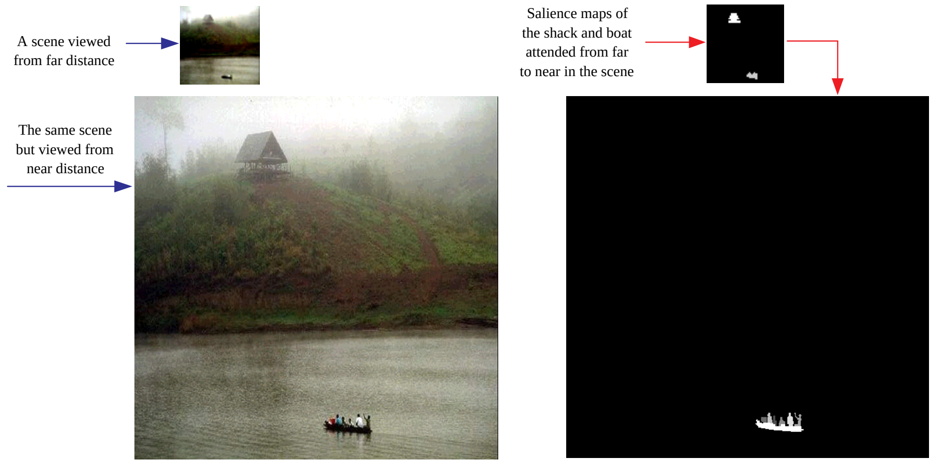

Fig. 5. An outdoor scene photographed from far and near distance respectively. The obtained images shown here are the same scene but different resolutions. The salience maps are shown too and the grey scales indicate the different salience of the groupings.

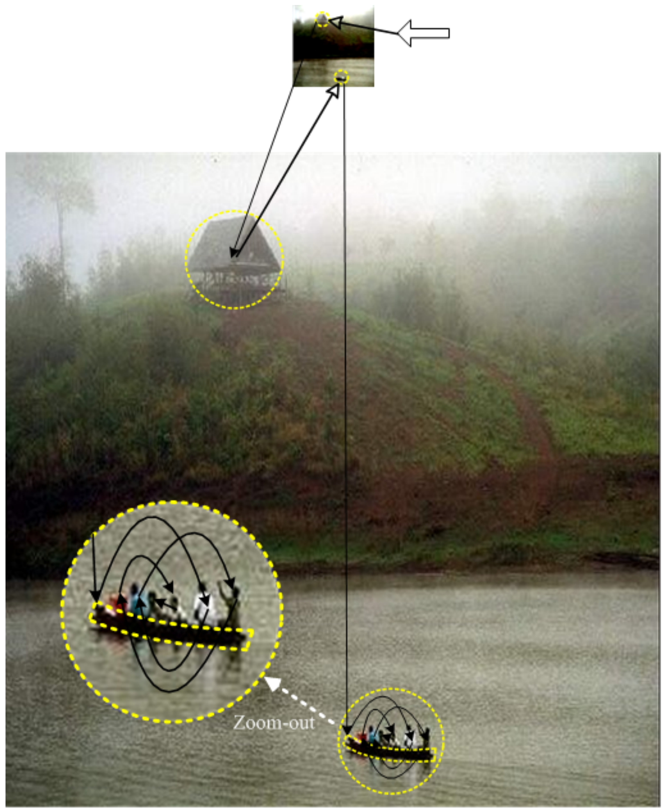

Fig. 6. The attention movements implemented for the outdoor scene: solid arrows indicate attentional movements at fine resolution and hollow arrows denote attention shifts at coarse resolution.

5 Conclusions and future research

Successful models of object-based attention require approaches different to the previous computable models of space-based attention (see [8] for a successful computable model of space-based attention). The new mechanisms must consider the selections of objects and groupings, without losing the advantages of space-based attention, such as selectivity by spatial locations and by feature. A good solution should integrate object-based and space-based attention together in a combined framework so that the attention model can work in a dynamic and natural environment. In consequence, multiple (such as features, spatial locations, objects, and groupings) and hierarchical selectivity can be implemented to deal with the complex visual tasks. The presented mechanism of hierarchical selectivity in our object-based attention model shows performance similar to human behaviour and also explores details in a manner useful for machine vision systems. Further research will extend the scope of top-down attention setting, for example, to allow enhanced and suppressed top-down control as well as more elaborate designation of whether it is "valuable" or not to check sub-groupings according to the current visual task.

References

1. S. Baluja, and D. Pomerleau, "Dynamic relevance: Vision-based focus of attention using artificial neural networks," *Artificial Intelligence*, 97, pp. 381-395, 1997.
2. M. Behrmann, R. S. Zemel, and M. C. Mozer, "Occlusion, symmetry, and object-based attention: reply to Saiki (2000)," *Journal of Experimental Psychology: Human Perception and Performance*, 26(4), pp. 1497-1505, 2000.
3. C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuity," *Human Neurobiology*, 4:481-484, 1985.
4. R. Desimone, and J. Duncan, "Neural mechanisms of selective visual attention," *Ann. Rev. Neurosci.*, 18, pp. 193-222, 1995.
5. R. Desimone, "Visual attention mediated by biased competition in extrastriate visual cortex," *Phil. Trans. R. Soc. Lond. B*, 353, pp. 1245-1255, 1998.
6. J. Duncan, "Converging levels of analysis in the cognitive neuroscience of visual attention," *Phil. Trans. R. Soc. Lond. B.*, 353, pp. 1307-1317, 1998.
7. H. Greenspan, S. Belongie, R. Goodman, P. Persona, S. Rakshit, and C. H. Anderson, "Overcomplete steerable pyramid filters and rotation invariance," In proc. IEEE Computer Vision and Pattern Recognition, pp. 222-228, Seattle, Washington, 1994.
8. L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11), pp. 1254-1259, 1998.
9. L. Itti and C. Koch, "Salience-Based serach mechanism for overt and covert shifts of visual attention," 40(10-12):1489-1506, 2000.
10. S. Kastner and L. G. Ungerleider, "Mechanisms of visual attention in the human cortex," *Annu. Rev. Neurosci.*, 23:315-341, 2002.
11. S. E. Palmer, *Vision Science-Photons to Phenomenology*, Cambridge, MA: MIT Press, 1999.
12. B. J. Scholl, "Objects and attention: the state of the art," *Cognition*, 80, pp. 1-46, 2001.
13. J. K. Tsotsos, et al. "Modelling visual attention via selective tuning," *Artificial Intelligence*, 78, pp. 507-545, 1995.
14. Yaoru Sun and Robert Fisher, "Object-based visual attention for computer vision," submitted to *Artificial Intelligence*.