(51) **International Patent Classification:**
G06T 5/00 (2006.01)

(21) **International Application Number:**
PCT/EP2007/054195

(22) **International Filing Date:** 27 April 2007 (27.04.2007)

(25) **Filing Language:** English

(26) **Publication Language:** English

(30) **Priority Data:**
06300418.8   28 April 2006 (28.04.2006)   EP
06300538.3   31 May 2006 (31.05.2006)   EP

(71) **Applicant** *(for all designated States except US)*: **THOMSON LICENSING** [FR/FR]; Thomson Licensing, 46 Quai A. Le Gallo, F-92100 Boulogne-billancourt (FR).

(72) **Inventors; and**
(75) **Inventors/Applicants** *(for US only)*: **GU, XIAO DONG** [CN/CN]; Rm 03-09, 8F, Bldg A, Technology Fortune Center, 8 Xue Qing Rd, Hai Dian District, Beijing 100085 (CN). **CHEN, ZHI BO** [CN/CN]; Rm 03-09, 8F, Bldg A, Technology Fortune Center, 8 Xue Qing Rd, Hai Dian District, Beijing 100085 (CN). **CHEN, QU QING** [CN/CN]; Rm 03-09, 8F, Bldg A, Technology Fortune Center, 8 Xue Qing Rd, Hai Dian District, Beijing 100085 (CN).

(74) **Agent: ROSSMANITH, Manfred**; Deutsche Thomson OHG, European Patent Operations, Karl-Wiechert-Allee 74, 30625 Hannover (DE).

(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— with international search report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: SALIENCE ESTIMATION FOR OBJECT-BASED VISUAL ATTENTION MODEL
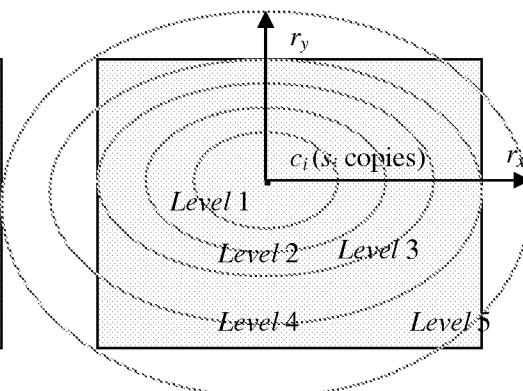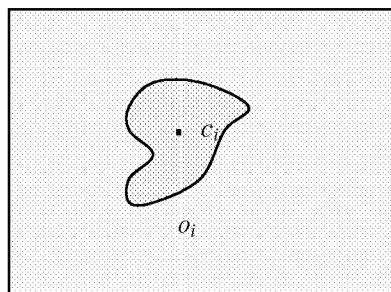
(57) **Abstract:** The present invention provides a salience estimation method for object-based visual attention model. The method comprises steps of segmenting the image into a plurality of objects to be estimated, extracting feature maps for each segmented object, calculating the saliences of each segmented object in a set of circles defined around a centre pixel of the object based on the extracted feature maps, and integrating the saliences of each segmented object in all circles in order to achieve an overall salience estimation for each segmented object. The present invention is much more human vision inosculated and of low computing complexity.

SALIENCE ESTIMATION FOR OBJECT-BASED VISUAL ATTENTION
MODEL

FIELD OF THE INVENTION

The present invention relates to a method for estimating the salience of an image, and more particularly to a salience estimation method for object-based visual attention model.

BACKGROUND OF THE INVENTION

As a neurobiological conception, attention implies the concentration of mental powers upon an object by close or careful observation. Attention area is the area in a picture where tends to catch more human attention. The system designed to automatically detect the attention area of a picture is called attention model. The detected attention area is widely utilized in many kinds of applications, such as accumulating limited resource in an attention area, directing retrieval/search, simplifying analysis, etc.

Fig.1 indicates the general architecture of a mostly used attention model. First, an image to be estimated is inputted into the attention model. Then the feature of intensity, colour, orientation, etc. will be achieved after the step of feature extraction. In the third step the salience of said features are estimated. After the steps of fusion scheme and post-processing the attention area is finally got.

Different from attention models used in most previous machine vision systems which drive attention based on the spatial location hypothesis with macro-block (MB) being the basic unit, other models which direct

visual attention are object-driven, called object-based visual attention model.

A lot of researches on MB (macro-block) spatial-based visual attention are established as proposed by L.Itti et al., "A Model of Salience-Based Visual Attention for Rapid Scene Analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 20, No.11, November 1998 and by Y.F.Ma et al., "A User Attention Model for Video Summarization", ACM Multimedia'02, pp.533-542, December 2002. However, object-based visual attention is not so widely studied because of its inherent difficulty. Y.Sun et al. propose a framework of object-based visual attention in "Object-based Visual Attention for Computer Vision", Artificial Intelligence, pp.77-123, May 2003. Another object-based visual attention model is presented by F.Orabona et al., "Object-based Visual Attention: a Model for a Behaving Robot", 3rd International Workshop on Attention and Performance in Computational Vision, June 2005. Both object-based visual attention schemes still follow the general architecture of attention model listed in Fig.1. All the processes except "salience estimation" are directly inherited from Itti's MB spatial-based visual attention model.

No matter in MB spatial-based or in object-based visual attention models, low level spatial/temporal features are first extracted, and then for each salient (different, outstanding from its surroundings; or say, more attractive) feature map of each unit is estimated over the whole picture, after that a master "salience map" is generated by feeding all feature maps in a purely bottom-up manner.

Compared with object-based visual attention model, the MB spatial-based visual attention model is a much easier and faster creation. However, it has several

5   inherent disadvantages:

1) The attention area breaks natural object boundary;

2) Each micro-block may cover lots of natural objects.

10   So, the extracted feature of the micro-block is a mixed property of all these natural objects and thus will lower down attention area detection precision.

The key issue of the object-based visual attention

15   model lies in two aspects: one is the object grouping before feature extraction, the other is the particular efficient salience estimation of each object over all the objects in the image. The central idea of the currently used salience estimation scheme is based on Gauss

20   distance measure as presented by Y.Sun et al.

Denote x as the object to be salience estimated, y_i(i=1,2,…,n) as all the background objects, w as the maximum of the width and height of the input image, and $\|x-$

25   $y_i\|$ as the physical distance between x and y_i , so the Gauss distance is defined as the formula(1),

$$d_{gauss}(x,y_i) = \left(1 - \frac{\|x-y_i\|}{w-1}\right)^{e^{-\frac{1}{2\sigma^2}\|x-y_i\|^2}}$$

(1)

with the scale σ set to w/ρ, where ρ is a positive integer and generally $1/\rho$ may be set to a percentage of w

30   such as 2%, 4%, 5% or 20%, 25%, 50%, etc.

Denote $S_F(x, y_i)$ as the absolute difference of object x and y_i in feature F, then the salience estimation $S_F(x)$ as the overall salience degree of object x in feature F can be expressed as Formula (2).

$$S_F(x) = \frac{\sum_{i=1}^{n} S_F(x, y_i) \cdot d_{gauss}(x, y_i)}{\sum_{i=1}^{n} d_{gauss}(x, y_i)}$$

(2)

By the definition of the salience estimation, it can be concluded that:

1. The larger difference between the object and its surroundings exists, the more salient the object is.

2. The closer the object and its feature differed surroundings is, the more salient the object is. That is, human vision decreases its ability to distinguish the difference according to distance. The attenuation coefficient is measured by d_{gauss}, which is coherent with the visual physiology thesis.

This guarantees $S_F(x)$ is a useful salient estimation in feature F. Unfortunately, some important human perception properties are not considered in $S_F(x)$.

Fig.2a is an original image of Skating to be estimated and Fig.3a is the salience estimation result of Fig.2a using the conventional object-based visual attention model.

Fig.2b is an original image of Coastguard to be estimated and Fig.3b is the salience estimation result of Fig.2b using the conventional object-based visual attention model.

Both in Fig.3a and Fig.3b, white colour means a very outstanding object while black colour means not

salient one, the grey level between white and black represents the salience degree.

From Fig.3a we can see that the audience is considered salient, because its colour greatly differs from its neighbour's, but actually the part of audience contains no details. Viewers usually will not focus on the audience and recognize it as "video texture".

Also in Fig.3a, there is a little grey block on the left of the female dancer's head. The block consists of a piece of white skating rink which is circled by black male clothing and female skin, and it is salient in this local area. But when all comes to all, this block is a part of the large skating rink and will not attract viewers' attention. This is called "Local effect". Because of the local effect, the accumulated difference between the object and its neighbours is large and thus it is recognized as "salience".

From forgoing description we can see that the conventional object-based visual attention model is not efficient enough and there are a lot of human vision properties not considered:

1. Object size – The estimation of the influence that the object size on salience degree is a complex problem. For example, (a) if all neighbouring objects y_i are of the same size s and the size of object x decreases from s to 0, as a result the salience degree of x ($S_F(x)$) will decrease gradually; (b) if all neighbouring objects y_i are of the same size s and the size of object x decreases from s_1 to s_2 ($s_1 \gg s$, and $s_1 > s_2 > s$), $S_F(x)$ will increase gradually. Thus we know that the relationship

between object size and salience degree is not monotonous. And the problem becomes even more complex when each of the objects may have an arbitrary size.

2. Local effect – If an object is not salient among its near neighbours (local area) while the far neighbours are greatly different from the object, there are two possible results: (a) the object is not salient at all inside the whole image; (b) the local area as a whole is salient inside the image with the object being a member of the local area. No matter in which case, the salient degree of the object does not match what defined above.

3. Video texture – Suppose the object features of an image are uniformly random, human will usually ignore the details of the whole image and not any object of the image is salient, while the above defined $S_F(x)$ will be a large number for any of the objects in the image.

With all these limitations, the conventional object-based visual attention model is far from applicable. Therefore an improved object-based visual attention model is desirable.

SUMMARY OF THE INVENTION

The present invention provides a salience estimation scheme for object-based visual attention model employing a multi-level concentric circled scheme capable of lowering the computing complexity and being more applicable.

In one aspect, the invention provides a method for estimating the salience of an image. It comprises steps of segmenting the image into a plurality of objects to be

estimated; extracting feature maps for each segmented object; calculating the saliences of each segmented object in a set of circles defined around a centre pixel of the object based on the extracted feature maps; and

5   integrating the saliences of each segmented object in the all circles in order to achieve an overall salience estimation for each segmented object. According to one preferred embodiment, the step of extracting feature maps is based on the measure of image colour variation.

10   According to another preferred embodiment, the step of calculating the salience of each segmented object comprises a sub-step of comparing colour features of the object to be estimated with that of any other object in each circle defined around the object to be estimated.

15

Advantageously, the object-based visual attention model based on multi-level concentric circled salience estimation scheme of the present invention presents an efficient framework to construct object-based visual

20   attention model, which is of low computing complexity and much more human vision inosculated.

Other characteristics and advantages of the invention will be apparent through the description of a

25   non-limiting embodiment of the invention, which will be illustrated with the help of the accompanying drawings.

BRIEF DESCRIPTION OF DRAWINGS
Fig.1 illustrates a general architecture of a

30   mostly used attention model;
Fig.2a illustrates an original image of Skating to be salience estimated;
Fig.2b illustrates an original image of Coastguard to be salience estimated;

Fig.3a is the salience estimation result of Fig.2a using the conventional object-based visual attention model;

Fig.3b is the salience estimation result of Fig.2b using the conventional object-based visual attention model;

Fig.4 illustrates the multi-level concentric circled scheme of the salience estimation according to a preferred embodiment of the present invention;

Fig.5 illustrates an example definition of *texture(.)* in the invention;

Fig.6a is an example of segmentation result of Fig.2a according to the preferred embodiment of the present invention;

Fig.6b is another example of segmentation result of Fig.2b according to the preferred embodiment of the present invention;

Fig.7a illustrates the estimated salience result of Fig.2a using the salience estimation scheme according to the preferred embodiment of the present invention; and

Fig.7b illustrates the estimated salience result of Fig.2b using the salience estimation scheme according to a preferred embodiment of the present invention.

DETAIL DESCRIPTION OF PREFERRED EMBODIMENTS

The technical features of the present invention will be described further with reference to the embodiments. The embodiments are only preferable examples without limiting to the present invention. It will be well understood by the following detail description in conjunction with the accompanying drawings.

From the foregoing description we can see that the salience estimation process can be denoted as:

Input: An image $I=\{p_i|i=1...w*h\}$ with width w and height h;

Output: Salience map $sal[1...w*h]$ with $sal[i]$ the salient degree of pixel p_i in the image.

5

The method of the present invention mainly includes three steps described as below:

Step 1 – Pre-processing (image segmentation)

10

An image I is first decomposed into a set of objects $I=\{o_1, o_2, ..., o_n\}$ in this step. Image-based segmentation and grouping play a powerful role in human vision perception, a lot of researches have been developed in this area. In this invention we adopt the object segmentation scheme proposed by P.F.Felzenszwalb et al., "Image Segmentation Using Local Variation", IEEE Computer Society on Computer Vision and Pattern Recognition, Jun.1998, which is based on measures of image colour variation. Here gives a simple description of this scheme.

Before processing, an undirected graph $H=(V, E)$ is defined based on the image I, where each pixel p_i of I has a corresponding vertex $v_i \in V$, and an edge $(v_i, v_j) \in E$ connects vertices v_i and v_j. The precise definition of which pixels are connected by edges in E depends on the expression (1-1).

$$E = \left\{(v_i, v_j) \mid \left\| p_i - p_j \right\| \le d \right\}$$

(1-1)

30 For some given distance d., a weight function on the edges, $weight(.)$, provides some non-negative measures of similarity (or difference) between the individual vertices v_i and v_j. Define $weight(.)$ as expression (1-2),

$$weight((v_i,v_j)) = \begin{cases} \left|Color(v_i) - Color(v_j)\right| & (v_i,v_j) \in E \\ \infty & otherwise \end{cases} \quad (1\text{-}2)$$

where $colour(v_i)$ is the colour of pixel p_i in the image.

$S=\{C_i\}$ denotes a segmentation of V and each C_i corresponds to a segmented object. Define the internal variation of C as formula (1-3),

$$Int(C) = \max_{e \in MST(C,E)} weight(e) \quad (1\text{-}3)$$

where $MST(C, E)$ is a minimum spanning tree of C with respect to the set of E.

Define the external variation of two objects C_1 and C_2 as formula (1-4).

$$Ext(C_1,C_2) = \min_{v_i \in C_1, v_j \in C_2} weight((v_i,v_j)) \quad (1\text{-}4)$$

The process of the segmentation is to make the expression (1-5) satisfied for any two of the segmented objects:

$$Ext(C_1,C_2) \le \min\left(Int(C_1) + k/\left|C_1\right|, Int(C_2) + k/\left|C_2\right|\right) \quad (1\text{-}5)$$

where k is a constant number set to 100 in our implementation. Denote $Int(C)+k/|C|$ as the extended internal variation inside object C.

To achieve the segmented result, E is first sorted into $\pi=(e_1, e_2,...,e_m)$ by non-decreasing edge weight and initially segment the image into $w*h$ single pixel objects, then for each $e_q=(v_i,v_j)$ $(q=1,2,...,m)$ repeat the next process: if v_i and v_j belong to different objects and $weight(v_i,v_j)$ is not larger than the extended internal variation$(Int(C)+k/|C|)$ of the two objects they belong to, the two objects are merged to form a new single object.

It can be seen, this gives an efficient object segmentation scheme which will not cost too much of computing resource. In implementation, here uses an 8-connected neighbourhood for constructing E, that is $d=1$. Fig.6a and Fig.6b provide the segmentation results of Fig.2a and Fig.2b respectively.

Step 2 – Pre-processing (feature extraction)

With y_i, r_i, g_i, b_i denoting the luminance, red, green and blue channels of pixel p_i, following we extract the features of each object segmented in step 1.

Considering the definition of extended internal variation $Int(C)+k/|C|$, wherein $k/|C|$ is an addition to the internal variation because underestimating the internal variation is bad for preventing components from growing. As a result, small objects are more likely to grow regardless of the internal variation inside. For example, in the Skating example of Fig.6a, the largest part of the black clothing mainly consists of black pixels but still there are some exceptions of white pixels.

To solve the feature extraction problem in the mentioned situation, an operator $Major(f, o)$ is defined in the feature map $F=f(v_i)$ of an object $o=\{v_1, v_2, …, v_t\}$. The returned value of $Major$ (f, o) is the representative feature of the object o which is defined to satisfy (d_1, d_2 and η are constant number, set to 2, 64 and 95% respectively in our implementation):

(1) If there exists a range [min, max] meeting the expression $max-min+1 \leq d_1$ and the percentage of features whose values inside the range [min, max] over the whole

feature map $F=\{f(v_1),\ f(v_2),\ ...,\ f(v_t)\}$ are not smaller than η, *Major(f, o)* is then defined as the average value of those features whose values are inside the range [*min, max*].

(2) Otherwise, if an object size is larger than a constant d_2, the object is divided into two sub-objects by the same process as in step 1 and then loop above step for each sub-object; otherwise if the object is too small, just define *Major(f, o)* as the average value of all the features.

With operator *Major(f, o)*, Y_i, R_i, G_i, B_i and Y_i being defined as the luminance, red, green, blue and yellow channels of object o_i (negative values are set to 0):

$$Y_i = Major(y, o_i)$$

$$R_i = Major(r, o_i) - \big(Major(g, o_i) + Major(b, o_i)\big)/2$$

$$G_i = Major(g, o_i) - \big(Major(r, o_i) + Major(b, o_i)\big)/2$$

$$B_i = Major(b, o_i) - \big(Major(r, o_i) + Major(g, o_i)\big)/2$$

$$Y_i = \big(Major(r, o_i) + Major(g, o_i)\big)/2 - \big|Major(r, o_i) - Major(g, o_i)\big|/2 - Major(b, o_i)$$.

The intensity feature is extracted as formula (2-1).

$$I_i = Y_i \qquad\qquad (2-1)$$

The "colour double-opponent" based colour features are extracted as formula (2-2) and (2-3).

$$RG_i = R_i - G_i \qquad\qquad (2-2)$$

$$BY_i = B_i - Y_i \qquad\qquad (2-3)$$

Orientation will be a certain complex feature in object based visual attention model. Since all the objects are segmented according to colour variations, the object itself then will not contain any orientation information except the border of the object. Because of

this special property of the segmented objects, we will
not consider orientation in the implementation.

5   Comparing with orientation, motion will be a more
possible additional feature since currently optical flow
techniques become more and more mature.

But for simplicity, we only consider the three
feature maps I_i, RG_i and BY_i in present invention.
10

Step 3 - Salience Estimation

After the above two steps the image I is segmented
into objects $I=(o_1, o_2, …, o_n)$, and three feature maps I_i,
15   RG_i and BY_i ($i=1…n$) are extracted. The remaining problem
is how to estimate the salience map for each feature map
F ($F\in\{I, RG, BY\}$), denoted as $Sal_F(o_i)$.

For any object o_i of the image, denote s_i as the
20   size (the number of pixels inside the object) and $c_i=(X_i,$
$Y_i)$ as the centre pixel of the object. X_i and Y_i are
described as formula (3-1).

$$X_i = \frac{1}{s_i}\sum_{j=1}^{s_i} x_j \quad ; \quad Y_i = \frac{1}{s_i}\sum_{j=1}^{s_i} y_j \qquad (3\text{-}1)$$

25   During the salience estimation process, each pixel
of the o_i is indistinctively considered equal to the
center pixel c_i, so the object is considered duplicated s_i
copies of the center pixel as shown in Fig.4.

30   Based on this assumption, there presents a multi-
level concentric circled scheme for salience estimation
of o_i. In the first step of this scheme, there defines a
set of concentric circles circled around the center pixel

c_i of the object, $C_1...C_t$ (C_j is an ellipse with horizontal radius r_x and vertical radius r_y, and is called Level j circle) are distributed from the near neighbouring areas of the center pixel c_i to the far neighbouring areas. For each level j circle, estimate the salience of o_i inside C_j, denoted as $SalC_F(o_i, C_j)$, and the overall estimated salience of o_i is then defined as formula (3-2), where k_t is a constant number for linear integration.

$$Sal_F(o_i) = \sum_{j=1}^{t} k_t \cdot SalC_F(o_i, C_j)$$

(3-2)

Then, given an area C_j and an object o_i in C_j with feature F extracted over C_j, how to estimate the salience of o_i inside C_j considering human vision properties? Here we first give the definition of the operation $SalC_F$:

(1) Set S as the set of objects with center pixel inside C_j.

(2) For each object o_u in S, define $F'_u = abs(F_u - F_i)$. Then use the follow formula to calculate the weighted average of F'_u in S.

$$avgF' = \frac{\sum_{\forall u, c_u \in S} F'_u \cdot s_u}{\sum_{\forall u, c_u \in S} s_u}$$

(3-3)

(3) Define ρ as the percentage of pixels in S with F'_u not larger than $avgF'$:

$$\rho = \frac{\sum_{\forall u, c_u \in S} bool(F'_u \leq avgF') \cdot s_u}{\sum_{\forall u, c_u \in S} s_u}$$

(3-4)

where $bool(exp)$ returns 1 when exp is a true determinant else returns 0.

(4) With the definition of a detection function $texture(.)$ as shown in Fig.5, $SalC_F$ is set as below,

$$SalC_F(o_i, C_j) = avgF' \times texture(\rho)$$

(3-5)

Where $texture(.)$ is an empirical function of ρ for detection of "audience area", i.e. the area with random

featured objects such as audience, which is more expected
not to be recognized as attention. The detection function
texture (ρ) satisfies that the lower the value of ρ is,
the bigger the value of *texture* (ρ) will be, and thus the
5   more chance this area is recognized as an "audience area"
i.e. the video texture of the image. By using this
detection function *texture* (.) there will be lower
probability that the non-attention objects in the area
are recognized as attention.
10

From the description above we can conclude the
salience estimation scheme as below:

(a) For each object o_i of the image, define a set
of concentric circles C_j (j=1...t).

15   (b) Calculate $SalC_F(o_i, C_j)$ according to above
definition of $SalC_F$.

(c) Integrate the salience estimation for all C_j
according to expression (3-2) to get the overall
estimated salience.
20

Fig.7a and Fig.7b respectively present the salience
estimation experimental results of Fig.2a and Fig.2b by
using the salience estimation scheme according to the
present invention. By using the present salience
25   estimation method, the audience in Fig.2a and the
background in Fig.2b are considered not salient, and the
little block on the left of the female dancer's head in
Fig.3a is also removed from Fig.7a. The present invention
is capable of handing the local effect and video texture
30   and it is more applicable.

Whilst there has been described in the forgoing
description preferred embodiments and aspects of the
present invention, it will be understood by those skilled

in the art that many variations in details of design or construction may be made without departing from the present invention. The present invention extends to all features disclosed both individually, and in all possible

5 permutations and combinations.

The present object-based visual attention model based on multi-level concentric circled salience estimation scheme gives a more accuracy on understanding

10 of the image and a far more computing efficiency, it has several advantages as below:

1. The invention presents an efficient framework to construct object-based visual attention model. It is of low computing complexity.

15 2. The presented framework is much more human vision inosculated. The un-considered human vision properties in conventional schemes (such as object size, local effect and video texture) are well issued.

3. The framework is extendable.

CLAIMS

1. A method for estimating the salience of an image, comprising steps of

-segmenting the image into a plurality of objects to be estimated;

-extracting feature maps for each segmented object;

characterized in that the method further includes steps of

-calculating the saliences of each segmented object in a set of circles defined around a centre pixel of the object based on the extracted feature maps; and

-integrating the saliences of each segmented object in the all circles in order to achieve an overall salience estimation for each segmented object.

2. The method according to claim 1, characterized in that the set of circles defined around the centre pixel of each segmented object are concentric circles.

3. The method according to claim 1, characterized in that the step of extracting feature maps is based on the measure of image colour variation.

4. The method according to any one of the preceding claims, characterized in that it further comprises a sub-step of comparing colour features of the object to be estimated with that of any other object in each circle defined around the object to be estimated in the step of calculating the salience of each segmented object.

5. The method according to claim 4, characterized in that said comparing step is implemented by calculating the absolute difference of colour feature between the

object to be estimated and any other object in each circle defined around the object to be estimated.

6. The method according to claim 1, characterized in that in the step of calculating the salience of each segmented object, a detection function is used for detecting the video texture of the image.

7. The method according to claim 1, characterized in that the overall estimated salience of an object is a linear combination of the saliences of the object in all circles.

Input Image

Feature Extraction

Intensity   Color   Orientation, etc.

Salience Estimation

Intensity   Color   Orientation
Saliency   Saliency   Saliency, etc.

Fusion Scheme

"master" Salient Map

Post-Processing

Attention Area
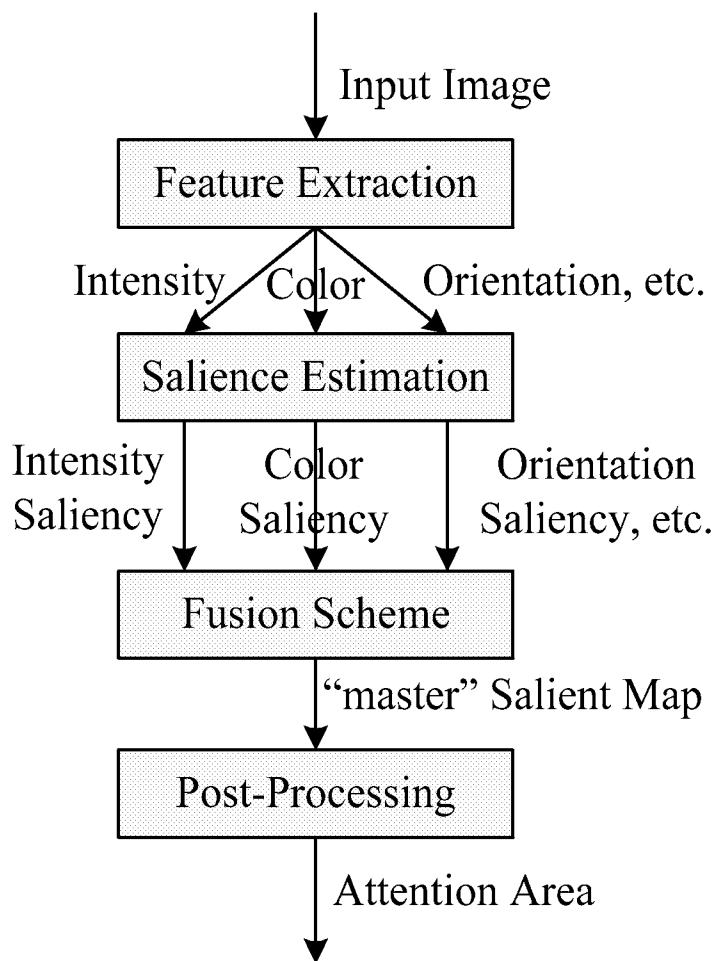
Fig.1
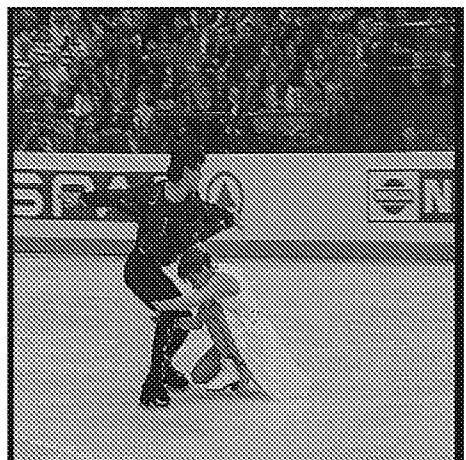
Fig.2a
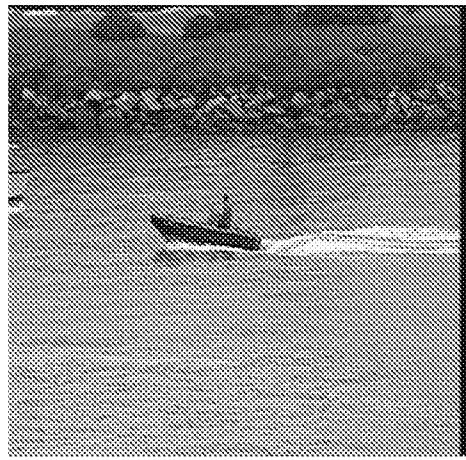
Fig.2b
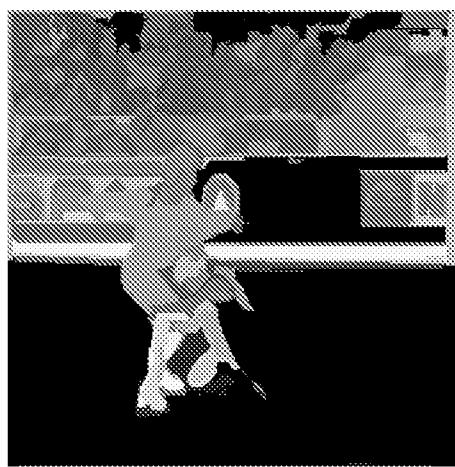
Fig.3a

Fig.3b
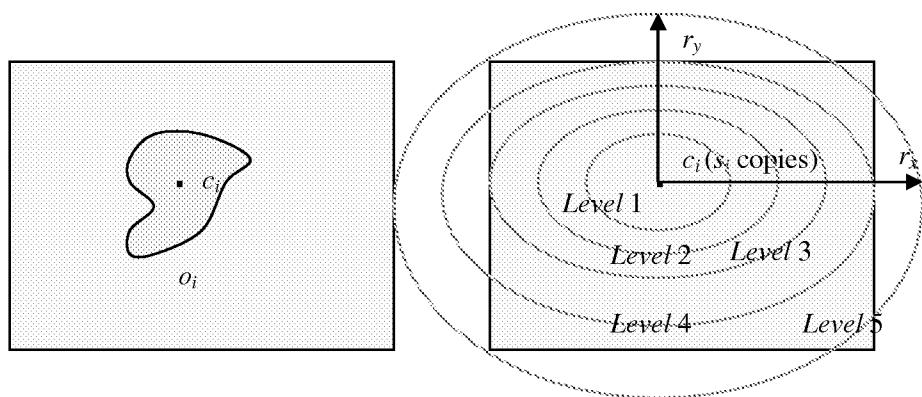
Fig.4
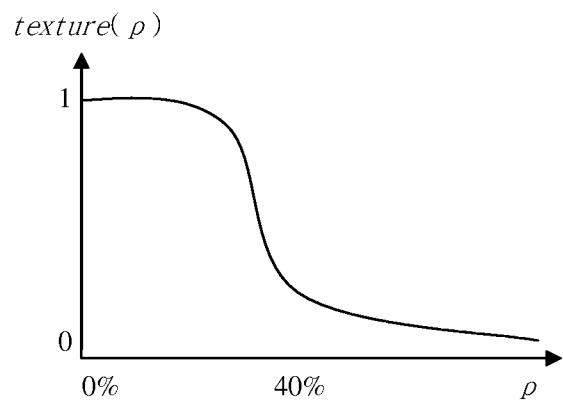
Fig.5
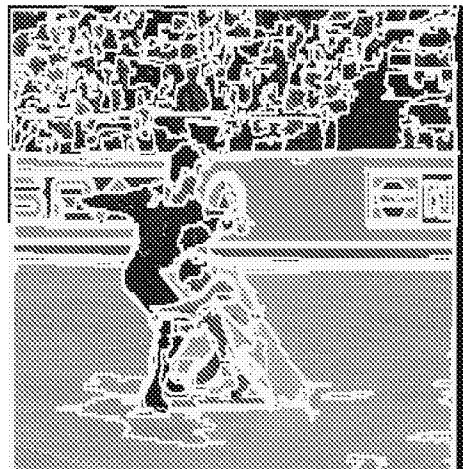
Fig.6a
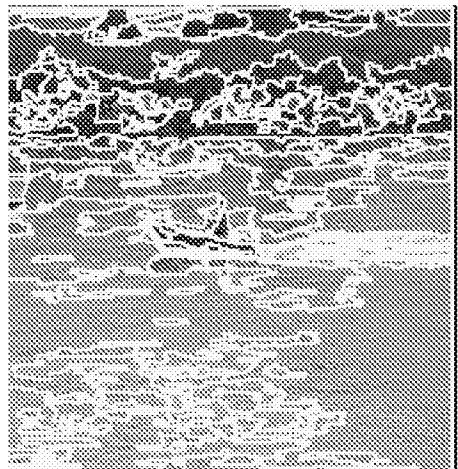
Fig.6b
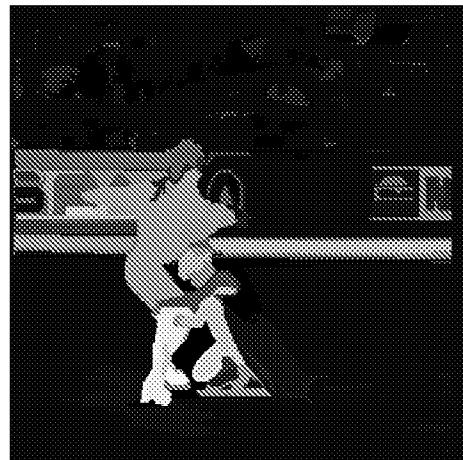
Fig.7a

Fig.7b

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06T5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06T

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, INSPEC, COMPENDEX

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LUO J ET AL: "On measuring low-level self and relative saliency in photographic images" PATTERN RECOGNITION LETTERS, NORTH-HOLLAND PUBL. AMSTERDAM, NL, vol. 22, no. 2, February 2001 (2001-02), pages 157-169, XP004315118 ISSN: 0167-8655 abstract page 159, right-hand column, lines 9-21 * sections "2.2. Weighting strategies", "3. Relative saliency features", "3.2. Color relative saliency features", "3.3. Texture relative saliency features" *	1-7

-/--

[X] Further documents are listed in the continuation of Box C.   [X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
25 June 2007	06/07/2007

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Eckert, Lars

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 1 017 019 A (EASTMAN KODAK COMPANY) 5 July 2000 (2000-07-05) abstract paragraphs [0020], [0029], [0031], [0035], [0038], [0040] - [0042]	1-6
A	ITTI L ET AL: "A SALIENCY-BASED SEARCH MECHANISM FOR OVERT AND COVERT SHIFTS OF VISUAL ATTENTION" VISION RESEARCH, PERGAMON PRESS, OXFORD, GB, vol. 40, no. 10-12, June 2000 (2000-06), pages 1489-1506, XP008060077 ISSN: 0042-6989 page 1490, right-hand column * section "2.1. Extraction of early visual features" * figure 1	1-7

International application No
PCT/EP2007/054195

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
EP 1017019	A	05-07-2000	JP	2000207564 A	28-07-2000
			US	6282317 B1	28-08-2001