Journal of Marine Science and Technology

Volume 24 | Issue 2 Article 9

MAXIMUM-LIKELIHOOD DIRECTIVITY FOR DEPICTING PERCEPTUAL CONTOURS

Chun-Shun Tseng
Department of Electrical Engineering, National Taiwan Ocean University,Keelung, Taiwan, R.O.C. Center of Excellence for the Oceans, National Taiwan Ocean University,Keelung, Taiwan, R.O.C.

Jung-Hua Wang
Department of Electrical Engineering, National Taiwan Ocean University, Keelung, Taiwan, R.O.C.

Ren-Jie Huang
Department of Electrical Engineering, National Taiwan Ocean University, Keelung, Taiwan, R.O.C.

MAXIMUM-LIKELIHOOD DIRECTIVITY FOR DEPICTING PERCEPTUAL CONTOURS

Acknowledgements

This work was supported by the National Science Council of Taiwan under grant NSC 102-2221-E-019-056.

Journal of Marine Science and Technology, Vol. 24, No. 2, pp. 152-162 (2016)
DOI: 10.6119/JMST-015-0616-2

MAXIMUM-LIKELIHOOD DIRECTIVITY FOR DEPICTING PERCEPTUAL CONTOURS

Chun-Shun Tseng[1, 2], Jung-Hua Wang[1], and Ren-Jie Huang[1]

Key words: Maximum-Likelihood Estimation, perceptual contours, directivity, entropy.

ABSTRACT

This paper presents a novel method based on Maximum-Likelihood Estimation (*MLE*) to evaluate pixel directivity for depicting image contours of objects as perceived by human eyes. The method is characterized by employing discrete masks with different shapes centered at a target pixel to sample gradient orientations of neighboring pixels for evaluating directivity of the target pixel, and applying *MLE* to determine one of these discrete sampling masks that best fits the orientation similarity of the target pixel. We show that such a fitting process in effect fulfils the similarity and proximity laws in *Gestalt* theory, and a salient alignment location can be determined by subjecting the optimal directivity in conjunction with the gradient magnitude of the target pixel to a Bayesian process. Finally, the directivity of salient alignment locations is incorporated with the extension field (Guy and Medioni, 1992) to detect perceptual contours. Experiments tested on complex images and underwater images are provided to justify the superiority of the work over others.

I. INTRODUCTION

Finding contours that fit human visual perception in acoustic/raster-scan images has long been an interesting research issue, and continuously considered one of the most desirable capability on machine vision in applications of security surveillance system, medical image analysis and autonomous inspection, etc. Recently, attentions have been drawn to autonomous underwater structures inspection (Foresti, 2001; Kim and Eustice, 2009; Galceran et al., 2014), such as ship hulls, pipelines and dams that need to be periodically inspected for purposes of maintenance, degradation assessment and underwater security, etc. Conventional inspection methods for underwater structures require either deploying human divers (Mittleman and Swan, 1993) or piloting a remotely operated vehicle (ROV) (Negahdaripour and Firoozfam, 2006). In modern underwater inspection, the use of remote devices to capture underwater images is constantly increasing, this is particularly true when it is necessary to avoid direct involvement of human divers operating in risky foul weather or radiation contaminated water. Compared to manual operations by divers, remote image capture and analysis are useful for performing subsequent inspections with improved coverage and better precision, and more importantly, for reducing overall need of human intervention. For instance, depiction of object contours in sonar images captured at the inspection site is never a trivial task for human operators, due to the low image quality mainly caused by granular effect. In addition, an inspector may easily lose his attention due to the lengthy time required for capturing quality images of every part of a target subject. Therefore, algorithms able to auto detect object contours in accord with the human visual perception are highly demanded.

Other applications concerning acoustic images can be the underwater object manipulation performed by a robotic vehicle equipped with acoustic cameras and robotic arms, and AUV (Autonomous Underwater Vehicle) navigation accomplished by using a sequence of landmarks, which are navigation cues fixed to the sea-bottom and can be observed by an imaging sonar. Planar array of sensors is an essential part of acoustic cameras to provide real-time three-dimensional (3D) maps of scenes that are some meters away from the planar array (Kunz and Singh, 2013; Mallios, 2014). These 3D maps are frequently displayed in a projective 2D version, like orthoscopic images or section images. Another kind of sonar system is the multibeam forward looking sonar (Quidu et al., 2012; Yufit and Maillard, 2013), which normally steers along the moving direction of the vehicle and provides real-time 2D acoustic images of the local sea-bottom presented in front of the vehicle. Due to the rapid development of acoustic imaging technology, many novel acoustic devices find their applications on-board of both ROVs and AUVs for a vast number of different tasks. Considering the increasing trend of autonomy in diverse engineering applications, the issue of detecting perceptual contours in acoustic images must be resolved.

In the case of the underwater raster-scan image, we nor-

Paper submitted 04/14/15; revised 05/24/15; accepted 06/16/15. Author for correspondence: Jung-Hua Wang (e-mail: jhwang@mail.ntou.edu.tw).
[1] Department of Electrical Engineering, National Taiwan Ocean University, Keelung, Taiwan, R.O.C.
[2] Center of Excellence for the Oceans, National Taiwan Ocean University, Keelung, Taiwan, R.O.C.

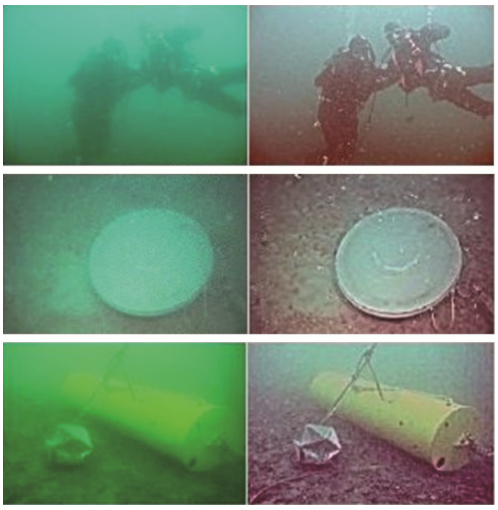


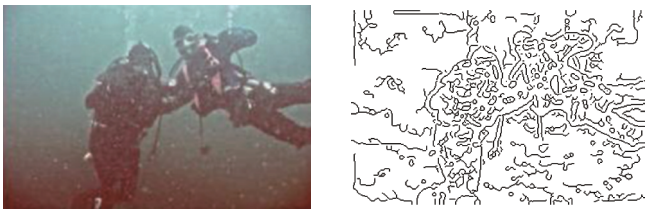
Fig. 1. Images before and after Bazeille et al.'s processing are shown in left and right columns, respectively (Courtesy of Bazeille et al., 2006).

mally have to consider its imaging characteristic different from in the air, namely physical limits such as light propagation in the water medium. Underwater raster-scan images are essentially characterized by their poor visibility because light is exponentially attenuated as it travels in the water, and hence the camera can only capture underwater scenes in a poor light condition. Statistically, light attenuation limits the visibility distance at about twenty meters in clear water and five meters or less in turbid water, and this limitation is mainly caused by absorption and scattering effects of the light in water. To be more specific, the forward scattering effect (i.e., randomly deviated light on its way from an object to the camera) generally leads to blurring of the image features, and the backward scattering (i.e., the fraction of the light reflected by the water towards the camera before it actually reaches the objects in the scene) generally limits the contrast of the images, thus generating a special "veil" that superimposes itself on the image and hides the scene. The absorption and scattering effects are arising from not only the water itself but other factors such as dissolved organic substances or small floating particles commonly known as "marine snow". In summary, the raster-scan images can suffer from the following problems: low contrast, blurring and noise. Fortunately, as shown in Fig. 1, (Bazeille et al., 2006) proposed a method able to alleviate the aforementioned problems by pre-processing raster-scan images to improve underwater image quality. Therefore, we can simply focus on solving the issue of perceptual contour detection without considering the low contrast and blurring problems by using Bazeille's method as a preliminary step.

As a common sense, our retina senses a pointillist array of light intensity values just like a digital camera recording a matrix of brightness values (Malik, 2006), but instead of simply recording the entire image pixel by pixel, the human visual system has the capability to interpret the pointillist array into essential information in various forms, including the most intriguingly visual perception in the brain. It is widely known that the human visual system has a preference for perceiving salient or aligned edges that form recognizable but not necessarily existing objects (Ullman, 1976), and when our eyes encountering a nature scene, what is actually perceived includes objects such as bridge, house or some particular alignments of boundaries, rather than a bunch of nuances of colors (Ellis, 1999). In other words, the human visual system has the capability to grasp the "gist" (i.e., alignment) vital for discriminating objects in the input images even if the existing contours of the objects are incomplete or vague (Sajda et al., 2010), for example, the contours of two divers in the top right image in Fig. 1 are in fact incomplete due to the inevitable presence of "marine snow", bubbles and noises when dealing with underwater images, yet our eyes can easily perceive the seemingly complete contours of two divers. Clearly, in order to develop machine executable algorithms that may find extensive applications in perceptual contour detection, decision-making and recognition, it is necessary to exploit, in view of computation perspective, some rationale principles underlying the human visual perception capability (Ren et al., 2008) in grasping and interpreting the "gist" such as aligned edges. To this end, this paper proposes an optimization method based on Maximum-Likelihood Estimation (*MLE*), which is capable of estimating the directivity (a value for representing the perceived alignment) of individual pixel useful for detecting salient alignment locations essential for constructing object contours consistent with human perception.

The proposed method mainly consists of three stages: (i) selecting pixels of interest; (ii) gradient orientations sampled with different shapes of a sampling mask $M(x, y)$ are statistically analyzed and their corresponding entropic values are ordered, and the concept of *MLE* is applied to the entropic results to determine the best shape of $M(x, y)$, which will be shown useful for selecting the optimal directivity value ($0 \leqq D(x, y) \leqq 1$) of a target pixel that essentially represents perception likelihood; (iii) subjecting $D(x, y)$ and the logarithmic gradient magnitude of the target pixel to a Bayesian process (Laplace, 1814) to determine whether the target pixel belongs to a salient alignment location, and finally, perceptual contours are depicted by incorporating the directivity of salient alignment locations with the concept of extension field (Guy and Medioni, 1992). The *MLE* process in stage (ii) are characterized in that (a) the shapes of a sampling mask are treated as a parameter set under estimation, so that directivity associated with pixels that best fit the orientation similarity of the target pixel can be estimated, meanwhile the well-known Gestalt laws (rules of understanding human visual ability to acquire and maintain meaningful perceptions) of proximity and similarity can thus be fulfilled (b) the problem of lacking directivity of individual pixel in region-based perceptual contour detectors (Desolneux et al., 2007; Von Gioi et al., 2010) can be overcome.

The remainder of this paper is organized as follows: literatures and researches related to the development of our work are discussed in the next section, and algorithmic details of



(a) (b)

Fig. 2. Images for demonstrating the human visual tendency. (a) A complex image. (b) Result of applying Canny detector to (a).

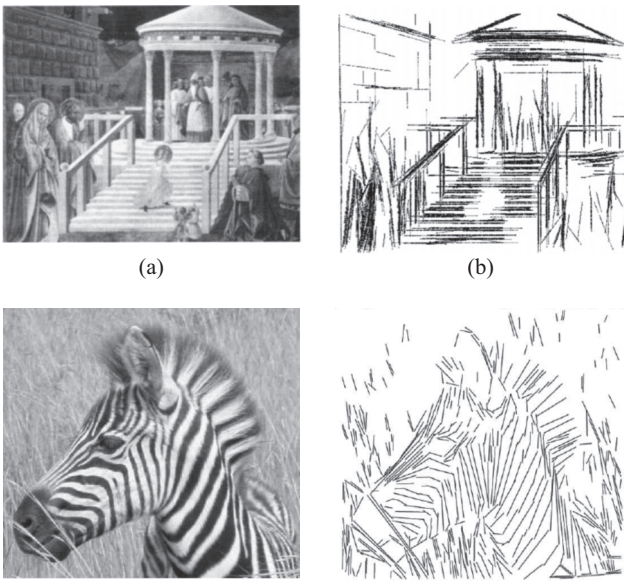
depicting perceptual contours are described in Section **III**. Characteristic analysis of the proposed method is given in Section **IV**, and extensive experimental results to justify the perceptual contour performance of our work are demonstrated in Section **V**. Finally, concluding remarks and future work are given in Section **VI**.

II. RELATED LITERATURES

The purpose of edge detection is to identify pixels in an input image at which the pixel intensity or brightness changes sharply such that most of object contours can be obtained from the detected edges. Ideally, edges composed of a set of straight and curved line segments useful for constructing object contours can be detected, with the least amount of data to be processed within the least amount of time and hence benefiting diverse research areas such as object segmentation (Cai and Miklavcic, 2013), pattern recognition (Shotton et al., 2008) and motion tracking (Cai et al., 2011; Sánchez-Nielsen, 2011), etc. To understand the motive of this work, suitability of using prior edge detection methods to facilitate perceptual contours depiction is examined and related approaches for depicting perceptual contours are also described.

1. Conventional Edge Detection Methods

Low-level edge detectors such as Sobel, Laplacian (Gonzalez and Woods, 2007), and Canny are based on differential calculation or criterion-based optimization (Canny, 1986). Although most edges, even the finest ones, can be detected by these methods with proper adjustment in parameters, it is very difficult, if not impossible, for them to produce perceptual contours. To see this, Fig. 2(a) shows an example in which most people would easily perceive two divers at a glance, rather than the overly connected or minute edges in Fig. 2(b), where the solid line and curves are results of the Canny detector. The intriguing tendency (Goldstein, 2013) to perceive the divers in the complex image of Fig. 2(a) reveals the fact that adjacent edges with an similar orientation (e.g. divers) are more noticeable to our eyes than the rest, thus enabling the human visual system to group these edges into perceptual contours, while unconsciously paying less attention to neighboring pixels of different orientations (e.g. those of divers' equipment). Thus, one can readily perceive the human body contour when looking at Fig. 2(a) even if the body is partially occluded by the "marine snow" and bubbles.



(a) (b) (c) (d)

Fig. 3. (a) (c) Test images. (b) (d) Results of applying Desolneux's method and Von Gioi's to (a) and (c), respectively.

Fig. 2(b) clearly indicates that using the Canny detector alone makes virtually any linking or grouping algorithms infeasible to construct contours, these contours are essential to enable a computing machine to "perceive" objects such as the divers in Fig. 2. In fact, a deeper insight into the human visual system, which is able to quickly compile complex scenes into simple object contours just to serve the survival purpose, would help substantiate the conjecture that accurate depictions of perceptual contours is not only essential in simplifying low-level operations of edge linking, but beneficial to high-level tasks of image analysis and pattern recognition. Thus, it is desirable to embody this capability of depicting perceptual contours on a computing device.

2. Approaches for Depicting Line Segment

Recognizing its great potentiality in engineering applications, numerous researchers have attempted to depict perceptual contours in various ways. Recently, gradient of pixels and Gestalt laws are considered in performing the contour depiction (Azriel and Thurston, 1971; Elder and Goldberg, 2002; Tseng et al., 2012). Challenging problems such as line scratch detection in old film can be overcome by using local statistical gradient estimation (Newson et al., 2014), and using Edge Orientation Histograms (EOH) as feature descriptors is always a popular way for recognition applications (Timotius and Setyawan, 2014). Desolneux et al. (2007) proposed a probabilistic method mainly based on gathering gradient orientations to depict aligned segments in an input image. The result of applying Desolneux's method to Fig. 3(a) is shown in Fig. 3(b), we can see that curves and junctions of clothes, creases and human body are not well depicted, even though the method

attempted to exploit similar orientation for continuity analysis on line segments.

Von Gioi et al. (2010) introduced a fast line segment detector (*LSD*) that incorporates Desolneux's method for evaluating gradient orientations within a line support region, that is, a growing region is used to sample and evaluate orientations in order to determine a principal orientation for the region. Despite the efficient computation owing to the region-growing scheme, only straight lines are preserved (as shown in Fig. 3(d)) when tested with real images such as the one in Fig. 3(c), accordingly reconstruction of curved contours of the zebra is virtually impossible. Both aforementioned methods are region-based as they are purposely designed to produce straight line segments each representing the principal orientation of a region, yet information other than this "lump sum" principal orientation, such as the directivity of individual pixel relative to its neighboring pixels is lacking. Therefore, an approach different from the region-based methods and capable of inferring psychologically important features such as curves and junctions is needed.

III. THE PROPOSED METHOD

This section explains in detail how to incorporate the concepts of *MLE* and entropy as the theoretic basis for evaluating the directivity.

1. Selecting Pixels of Interest

At the initial stage, a smoothed image I is obtained by applying a Gaussian blurring (McAndrew et al., 2010) with σ = 0.5 to the input image. Subsequently, gradients in two directions $\boldsymbol{G_h}$ and $\boldsymbol{G_y}$ are obtained by applying any gradient operations such as Sobel operators to I. Matrices of the magnitude $\boldsymbol{G}$ and orientation $\boldsymbol{\theta}$ are calculated as $\sqrt{\boldsymbol{G_h}^2 + \boldsymbol{G_y}^2}$ and $arctan(\boldsymbol{G_y}/\boldsymbol{G_h})$, respectively. $G^L(x, y)$ is then obtained by normalizing the result of taking a logarithm operation on $\boldsymbol{G}$, which is in fact a simple way to suppress the scattering range of the gradient magnitude. If $G^L(x, y)$ is smaller than 0.5, the corresponding value of $\boldsymbol{\theta}(x, y)$ is replaced with an angle value randomly chosen from [-90, 90], the rationale of such replacement lies in that pixels with $G^L < 0.5$ generally have diverse orientations, that is, we are not sure about which orientation is really associated with these pixels since their brightness don't change sharply. Considering the fact that partial perceptual contours themselves must be edges too, but the converse is not always true, thus it is obvious that a subset of all pixels in I can be selected as targets for the depiction of perceptual contours with less computation time. Furthermore, because the Canny detector itself includes aforesaid steps of Gaussian blurring and computing $\boldsymbol{G}$ and $\boldsymbol{\theta}$, unless otherwise stated, we will use the MATLAB-implemented Canny detector (McAndrew et al., 2010) hereinafter to select pixels of interest, and denote $G^L_{Can}(x, y)$ as the subsets of $G^L(x, y)$ containing gradient magnitudes of pixels detected by the Canny detector. Note that double thresholds 0.25 and 0.1 are used in Canny detector, as they are low enough to meet the requirement of preserving most of applicable edges.

2. Estimation of Directivity

At this stage, the semi-minor axis length *Rsr* and orientation of an elliptical sampling mask M^{Rsr} are allowed to change, just as in our previous work (Tseng et al., 2012). However, instead of using the lengthy iterative training process and changing the mask shape according to the entropy contained in the mask, here several different shapes of sampling mask are subjected to the *MLE* to determine the optimal shape of the mask for evaluating the orientation similarity of the target pixel. The following discussions explain one embodiment of it:

(i) Constructing an elliptic sampling mask M^{Rsr} centered on a selected pixel (x, y) by (1)

$$Rsj(x, y) = \frac{r^2}{Rsr(x, y)}, \tag{1}$$

without loss of generality, Rsr = 1, 2, 3 and 4 (pixels), *Rsj* and *Rsr* are the semi-major and semi-minor axis of an ellipse, respectively. The area of M^{Rsr} is defined by πr^2, and r = 4 is used throughout this paper.

(ii) $\boldsymbol{\theta}$ is quantized by $\left\| \frac{\theta \times Bn}{180} \right\| \times \frac{180}{Bn}$, Bn is the total number of bins and $\|x\|$ denotes the nearest integer to x. Note that M^{Rsr} is rotated to align its semi-minor axis with $\boldsymbol{\theta}(x, y)$ and remain un-rotated afterwards.

(iii) With the quantized $\boldsymbol{\theta}$, compute the histogram of orientation data within M^{Rsr}, where each bin (occurrence frequency) is denoted as $h_i^{Rsr}(x, y)$. In particular, we denote $h_{\max}^{Rsr}(x, y)$ as the largest bin among all bins $h_i^{Rsr}(x, y)$, i = 1, 2… Bn, and $h_T^{Rsr}(x, y)$ as the bin associated with the target pixel, respectively. The quantized orientation $h_{\max}^{Rsr}(x, y)$ in a sense corresponds to the principal orientation in the region-based methods (Desolneux et al., 2007; Von Gioi et al., 2010).

(iv) Defining the directivity D^{Rsr} for a selected pixel (x, y) as

$$\mathrm{D}^{Rsr}(x, y) = 1 - \frac{(EN^{Rsr}(x, y) + EN^{Rsr}_{offset}(x, y))}{EN^{Bn}_{\max}}, \text{where} \tag{2}$$

$$EN^{Rsr}(x, y) = -\sum_{i=1}^{Bn} h_i^{Rsr}(x, y) \log(h_i^{Rsr}(x, y)), \text{ and} \tag{3}$$

$$EN^{Rsr}_{offset}(x, y) = (1 - \alpha) \times (EN^{Rsr}_{\max}(x, y) - EN^{Rsr}(x, y)) \tag{4}$$

where $EN^{Bn}_{\max}$ and $EN^{\mathrm{Rsr}}_{\max}$ are the global maximum entropy and local maximum entropy defined in (11), re-

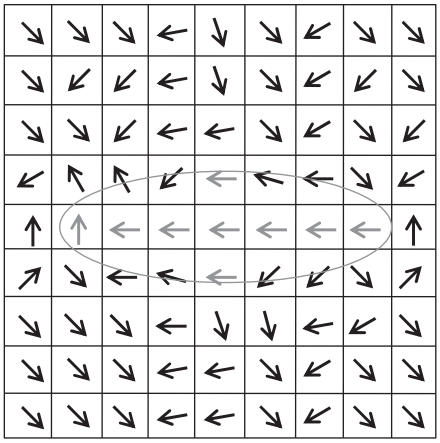


Fig. 4. Example for demonstrating a perceiving area *A* and a mask *M* (the gray region).

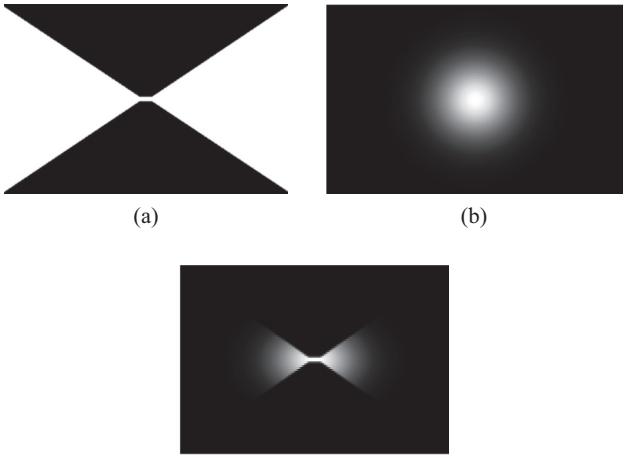


Fig. 5. (a) A 2D matrix *B*. (b) 2D Gaussian distribution (σ = 15). (c) Result of applying Hadamard product to (a) and (b).

spectively, and α is defined as $h_{\mathrm{T}}^{\mathrm{Rsr}}(x, y) / h_{\max}^{\mathrm{Rsr}}(x, y)$. The term (1-$\alpha$) mathematically accounts for the discrepancy between occurrence frequencies of the target pixel and the principal orientation, it offers interesting implications particularly related to human visual perception, details of which will be given later. Here, the base of the log( ) in (11) can use 2, e, or 10.

After executing the foregoing steps, we have obtained four separate directivity values D^{Rsr}. Next, the concept of *MLE* is applied to determine the optimal directivity representing the orientation similarity with the target pixel,

$$L(M^{Rsr}) = \prod_{i=1}^{Bn} f(a_i; M^{Rsr}), \tag{5}$$

where a_i can be treated as an observation or sample in the context of *MLE*, it corresponds to the occurrence frequency of a specific orientation within a perceiving area *A* of human vision (e.g., the orientation of arrows in Fig. 4). On the other hand, *M* can be viewed as an approximation for *A* (the gray region in Fig. 4) such that a_i can be treated as equivalent to h_i. Taking nature logarithm of (5) yields (6),

$$\ln L(M^{\mathrm{Rsr}}) = \sum_{i=1}^{Bn} \ln f(a_i; M^{\mathrm{Rsr}}) \tag{6}$$

where $\ln f(a_i; M^{Rsr})$ is written as $h_i^{\mathrm{Rsr}}(x, y) \log(h_i^{\mathrm{Rsr}}(x, y))$, which in fact corresponds to the entropy prescribed in (3). If we treat M^{Rsr} as a parameter set in *MLE*, then because the area of M^{Rsr} is fixed and the possible number of shapes of M^{Rsr} is finite in the discrete space, the theory of *MLE* requires that the maximum value of $L(M^{Rsr})$ must be corresponding to the optimal shape of M^{Rsr} for sampling (or observing) orientations in *A*. By the same token, we can simply use D^{Rsr} to approximate $\ln(L(M^{Rsr}))$, and the maximum of ordered $\boldsymbol{D}$ at (x, y), denoted as $D^O(x, y)$ hereafter, corresponds to the optimal directivity.

3. Determination of Salient Alignment Location

Examining Fig. 2(a) and (b) reveals the fact that not all detected edge pixels correspond to salient alignment locations. Given $D^O(x, y)$ and $G_{Can}^L(x, y)$, determination of salient alignment locations can be simply carried out in the context of Bayesian process, that is, if the parameter $G_{Can}^L(x, y)$ is treated as a *prior* probability of a target pixel being located on a salient alignment location, then the optimal directivity $D^O(x, y)$ can be treated as the *likelihood* of the target pixel being located on a salient alignment location, namely the *perception likelihood* $D^O(x, y)$ is a conditional probability of observing the events $\boldsymbol{h}^O(x, y)$ given that the target pixel is a perceptual one. Now, by plugging $G_{Can}^L(x, y)$ and $D^O(x, y)$ into the Bayesian formula in (7), the post probability of a target pixel being on a salient alignment location can be calculated as

$$S(x, y) = \frac{\mathrm{D}^{\mathrm{O}}(x, y) \times G_{Can}^L(x, y)}{\mathrm{D}^{\mathrm{O}}(x, y) \times G_{Can}^L(x, y) + [(1 - \mathrm{D}^{\mathrm{O}}(x, y)) \times (1 - G_{Can}^L(x, y))]} \tag{7}$$

Because the target pixel is either salient or not, the Bayesian decision can be simply made by the following rule: if $S(x, y) > 0.5$, the pixel at (x, y) is determined to be on a salient alignment location.

4. Depicting Perceptual Contours

In Guy's method (1992) an extension field *F* defined as

$$F = Ga \circ B \tag{8}$$

is used to depict perceptual contours, where *B* is a 150 × 100 butterfly pattern with binary values 0 and 1 (see Fig. 5(a)), *Ga* is a 2D Gaussian distribution with σ set to 15 (Fig. 5(b)).

Fig. 5(c) shows the resulting F by applying the Hadamard product of (8), where gradients in two directions $\boldsymbol{F_h}$ and $\boldsymbol{F_y}$ are obtained by applying any gradient operations such as Sobel operators to F. Matrices of the magnitude $\boldsymbol{F^G}$ and orientation $\boldsymbol{F^\theta}$ are calculated as $\sqrt{\boldsymbol{F_h^2}+\boldsymbol{F_y^2}}$ and $arctan(\boldsymbol{F_y}/\boldsymbol{F_h})$, respectively. As defined in Guy's method, each pixel **(x, y)** in the input image is associated with a voting matrix $[v^1, v^2, ..., v^{Bn}]$, and each element $F(x_f, y_f)$ (x_f = 1, 2, …, 150 and y_f = 1, 2, …, 100) carries a voting value (i.e., gradient magnitude $F^G(x_f, y_f)$**).** Also, the gradient orientation $F^\theta(x_f, y_f)$ of an element $F(x_f, y_f)$ is quantized and referred as an index j, j =1, 2, ..Bn, in accumulating the voting values $F^G(x_f, y_f)$ in v^j of the pixel located correspondingly at (x, y) in the input image. Furthermore, depending on the gradient orientation $\theta(x, y)$ of an interest pixel, F should rotate correspondingly in order to accumulate the votes due to the human visual tendency of noticing adjacent edges with an similar orientation. For instance, assume a vertical interest edge pixel (x, y) in the input image, because it actually has a gradient orientation of 0°, thus it is necessary to center F on (x, y) and rotate by 90 degrees. Afterwards, v^j (j = 1, 2, ..Bn) of each pixel (x, y) in the input image within the range correspondingly covered by F is accumulated by a voting value $F^G(x^f, y^f)$**.**

In this work, instead of using an identical F throughout as in Guy's method to accumulate voting values for each pixel, the directivity value $D^O(x, y)$ obtained in Section III is utilized to scale the size of F by multiplying $D^O(x, y)$ to the length and width of F. After subjecting all salient alignment locations one by one to this modified voting region F', each pixel in the input image can be characterized by a resulting matrix $V(x, y) = [v^1, v^2, ..., v^{Bn}]$ on the 50 orientations (as Bn = 50), and the corresponding index with the maximum value in $V(x, y)$ is defined as m. In order to avoid the nonlinear effect incurred by the boundary condition, namely when m equals v^1 and v^{Bn}, an extended resulting matrix V_{ex} is defined in (9), and a measure $V^O(x, y)$ more suitable for characterizing the degree of forming perceptual contours is prescribed in (10), where the index m is applied to (10), and L is denoted as the range for gathering voting results of similar orientations.

$$V_{ex} = [V, V, V]_{1\times 150}, \tag{9}$$

$$V^O(x, y) = \frac{1}{2L+1}\sum_{q=Bin+m-L}^{Bin+m+L} V_{ex}(q) \tag{10}$$

Note that, because the directivity at every salient alignment location is not necessarily identical, modifying the size of F into F' seems a rationale step approaching the human visual system in forming recognizable but not necessarily existing objects based on salient or aligned edges. In addition, the nonlinear calculation of V^O, which considers similar orientations while depicting object contours as the human vision does, can advantageously overcome the problem of unexpected saliency map strength values incurred by the linear moment calculation of vector weights (Guy and Medioni, 1992).

IV. CHARACTERISTIC ANALYSIS

To fully understand the proposed method, some elaborations are given below. In (3) the entropy $EN^{Rsr}(x, y)$ for each mask centered at the target pixel is calculated. With Q^{Rsr} denoting the total number of nonzero bins among all bins, the local maximum entropy $EN_{\max}^{\mathrm{Rsr}}$ is calculated by

$$EN_{\max}^{Rsr}(x, y) = -\log(\frac{1}{Q^{Rsr}}) \tag{11}$$

and the global maximum entropy $EN_{\max}^{Bn}$ yields when Q^{Rsr} = Bn, which is the extreme case when each pixel in M^{Rsr} by itself has a separate nonzero bin. With r = 4, the total number of pixels in M^{Rsr} is 50, which is the maximum number of bins that occurs when $Q^{Rsr} = Bn$.

Except for some special cases, $EN^{Rsr}(x, y)$ inversely stands for the orientation similarity associated with M^{Rsr}, i.e. a smaller $EN^{Rsr}(x, y)$ in general indicates a stronger similarity due to the distribution of $\boldsymbol{h}^{Rsr}(x, y)$ being more different from uniformity, and vice versa. Thus, using $\boldsymbol{h}^{Rsr}(x, y)$ to compute $EN^{Rsr}(x, y)$ allows us to conveniently infer the orientation similarity within M^{Rsr}, which in effect fulfils the similarity law of *Gestalt* theory (King and Wertheimer, 2005) that describes the perceptual tendency to group items (e.g. pixels, edges) into meaningful configurations if they are similar with respect to some features such as shape, colour or texture.

Human perception is a rather intriguing task in view of information theory, which could make simple measures such as $EN^{Rsr}(x, y)$ not sufficient, in some cases, for properly measuring the perceived directivity of the target pixel relative to the neighboring pixels. To see this, we first assume $EN_{offset}^{Rsr}(x,y) = 0$, then (2) is readily reduced to

$$D^{Rsr}(x, y) = 1 - \frac{EN^{Rsr}(x, y)}{EN_{\max}^{Bn}} \tag{12}$$

Next, we use Fig. 6 to show two examples with an infinite mask, in which a target pixel is enclosed by a dashed square, with symbols → and ↑ representing two different orientations 0° and 90°, respectively. One would easily perceive Fig. 6(a) as separate lines broken at the target pixel, and Fig. 6(b) as straight lines. Despite that Fig. 6(a) and (b) are perceived differently by the human visual system, entropy computation by (3) yields $EN^{Rsr}(x, y)$ = 0 for both cases, as $h_1^{Rsr}(x, y)$=1 (the occurrence frequencies of → in Fig. 6(a) and Fig. 6(b) are $\frac{\infty-3}{\infty}$ and $\frac{\infty}{\infty}$, respectively) and $h_2^{Rsr}(x, y)$ = 0 (the occurrence frequencies of ↑ in Fig. 6(a) and Fig. 6(b) are $\frac{3}{\infty}$ and

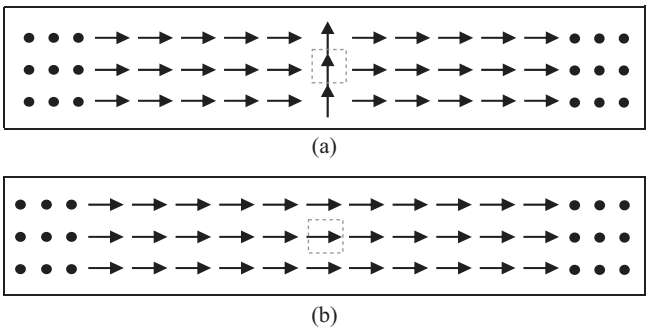


Fig. 6. (a) A target pixel is enclosed by a dashed square with a symbol ↑ representing 90°. (b) A target pixel is enclosed by a dashed square with a symbol → representing 0°.

$\frac{0}{\infty}$, respectively). Now plug $EN^{Rsr}(x, y) = 0$ into (12), the directivity values for Fig. 6(a) and Fig. 6(b) are both 1, which is obviously contradictory to the human visual perception. To avoid this contradiction, the target pixel "↑" in Fig. 6(a) should have a much smaller directivity value than the target pixel "→" in Fig. 6(b). Clearly, there exists a need of a compensation term in (12) to cope with this *perceptual contradiction effect* shown in Fig. 6. In this work, the compensation term EN^{Rsr}_{offset} is prescribed as in (4), in particular $(1-\alpha)$ is defined as the contradiction factor and it has two implications:

(i) EN^{Rsr}_{offset} is regulated by α, given $h^{Rsr}_{\max}(x, y) > h^{Rsr}_{T}(x, y)$ or even if $h^{Rsr}_{\max}(x, y)$ nearly equals one.

(ii) if $h^{Rsr}_{\max}(x, y)$ equals or close to $h^{Rsr}_{T}(x, y)$, then $\alpha \approx 0$, EN^{Rsr}_{offset} is virtually not needed.

To see the first implication, assume $h^{Rsr}_{\max}(x, y) \approx 1$ and $h^{Rsr}_{T}(x, y) \approx 0$, which corresponds to Fig. 6(a) where a majority of sampled data share the same orientation (i.e., 0°), namely a dominant mode exists and the entropy $EN^{Rsr}(x, y)$ is nearly zero. With the large contradiction factor $(1-\alpha) \approx 1$, the perceptual contradiction effect is very prominent, and a large value of EN^{Rsr}_{offset} is required to offset the contradiction effect. Now, by (4) $EN^{Rsr}_{offset} \approx EN^{Rsr}_{\max}$ due to $EN^{Rsr}(x, y) \approx 0$, thus a small directivity value of the target pixel can be correctly obtained by (2), and the problem in Fig. 6(a) is solved.

Next, it is interesting to see that as the number of ↑ increases in Fig. 6(a), the value of $(1-\alpha)$ decreases, and as the number of ↑ goes to infinite, the perceptual contradiction effect completely fades away, which is exactly equivalent to the situation in Fig. 6(b). Based on the above discussions, $EN^{Rsr}_{offset}(x,y)$ indeed can properly offset the perceptual contradiction effect, that is, with (4) a larger value of $EN^{Rsr}_{offset}(x,y)$ can be produced to correspondingly offset a stronger perception contradiction effect.

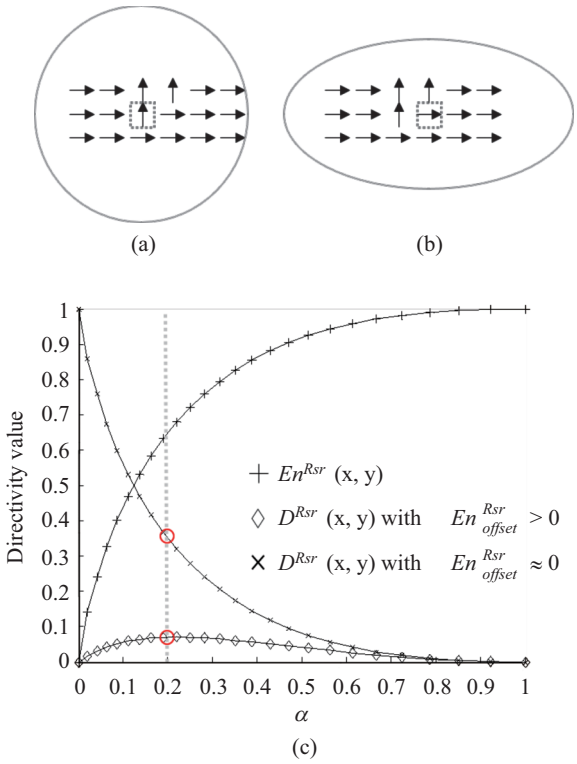


Fig. 7. (a) Example of large *Rsr*. (b) Example of small *Rsr*. (c) Directivity value vs. *α*.

As to the second implication, it simply states that either $h^{Rsr}_{T}(x, y)$ itself is the dominant mode (e.g., 0° in Fig. 6(b) and $h^{Rsr}_{T}(x, y)$ is the largest bin) or there are at least two major modes, i.e. $h^{Rsr}_{\max}(x, y) \approx h^{Rsr}_{T}(x, y)$, whichever the case is, the perception contradiction effect is insignificant, and EN^{Rsr}_{offset} is nearly zero, making (12) essentially equivalent to (2).

Another point worth noting: recall that $Rsr(x, y)$ is rotated to align with $\theta(x, y)$ in the first iteration and remain un-rotated, the four mask shapes are narrower to wider with $Rsr(x, y) = 1$, 2, 3 and 4, which not only facilitates pixels that share similar orientations with the target pixel to be covered as many as possible by M^{Rsr}, but also ensures pixels sharing no similar orientations with the target pixel to be encompassed for confirming the unlikely presence of directivity in M^{Rsr}. Thus, the optimal mask shape estimated by *MLE* covers a region containing pixels from which a directivity value that best characterises the target pixel can be evaluated. Fig. 7(a) shows an example of the mask with large *Rsr* centered at a target pixel of 90° (↑) when $\alpha = 0.2$ (three pixels of ↑ and fifteen pixels of →, i.e. $h^{Rsr}_{T}(x, y) = 3$ and $h^{Rsr}_{\max}(x, y) = 15$), whereas Fig. 7(b) shows an example of the optimal mask centered at a target pixel of 0° (→) when $\alpha \approx 1$ (i.e., $h^{Rsr}_{T}(x, y) = h^{Rsr}_{\max}(x, y) = 15$). By (2), the directivity value of the target pixel (enclosed

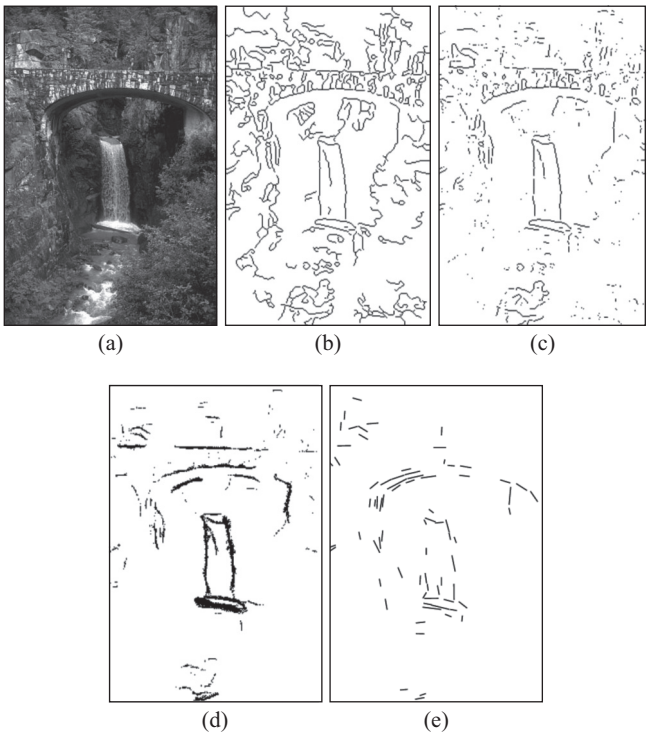


Fig. 8. (a) A complex image. (b) Detection result of applying Canny to (a). (c) Salient alignment locations determined from interest points (b). (d), (e) Contours depicted by applying our method and *LSD* to (a), respectively.

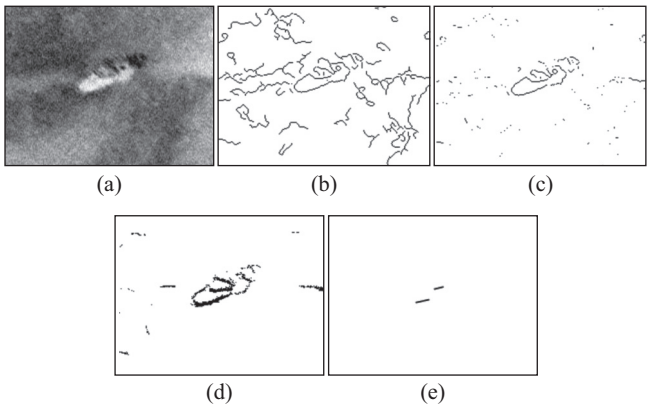


Fig. 9. (a) A shipwreck in a sonar image (Courtesy of Sture Hultquist www.abc.se/~m10354/uwa). (b) Detection result of applying Canny to (a). (c) Salient alignment locations determined from interest pixels (b). (d) (e) Contours depicted by applying our method and *LSD* to (a), respectively.

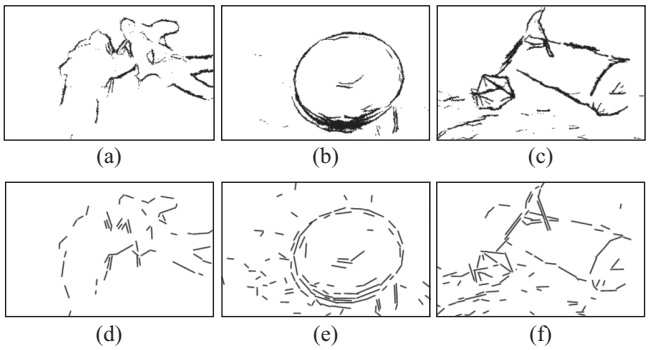


Fig. 10. (a), (b), (c) and (d), (e), (f) are detection results of applying our method and *LSD* to right column images in Fig. 1, respectively.

by a square) in Fig. 7(a) is 0.05, which is relatively smaller than the directivity value 0.36 of the target pixel in Fig. 7(b) by (12). As expected, the optimal mask shape Fig. 7(b) is more directional than that in Fig. 7(a). For viewing convenience, both directivity values 0.05 and 0.36 are marked as a red circle in Fig. 7(c).

We note that despite the histogram structures of Fig. 7(a) and (b) are the same, the mask shapes as well as the directivity values calculated with (2)-(4) for the two different pixels are quite different, sufficiently exhibiting the discriminant performance in line with our human visual perception. Statistical inferences of Fig. 7(a) and (b) are as follows: the different directivity values $D^{Rsr}(x, y)$ with different *Rsr* calculated according to (2)-(4) are indeed effective in measuring the directivity of a target pixel, verifying that the proposed entropy-driven scheme can implement the proximity law in *Gestalt* theory that describes the perceptual tendency to group items as meaningful configurations according to their nearness to one another (King and Wertheimer, 2005).

V. EXPERIMENTAL RESULTS

In this section, extensive experimental results are provided to justify the feasibility and effectiveness of our method. Note that if the standard deviation of $V(x, y)$ is smaller than 3, which means that there is no significant orientation in the voting matrix $V(x, y)$ and hence the corresponding $V^O(x, y)$ is set to zero. V^O is normalized to [0, 1] and thresholded by 0.2, and L is set to 5 in all the experiments of nature images. Although these parameters related to the extension field are set heuristically, the following test results of nature images reveal that our algorithm is not sensitive to these parameters. Binary images in Fig. 8(d) and (e) are shown first to compare the results of applying our method and *LSD* to Fig. 8(a), respectively. Even some edges of objects are missing in the detection result of Fig. 8(d), one can still distinctly perceive the bridge and waterfall in contrast that the detection result of *LSD* in Fig. 8(e) only roughly depicts an arch contour. In particular, comparing Fig. 8(c) with Fig. 8(b) verifies that noise interferences by leaves and textures with dissimilar orientations have been successfully suppressed. Fig. 9(d) is a ship contour obtained by applying our method to a sonar image (Fig. 9(a)), and Fig. 9(e) is the result of applying *LSD* to Fig. 9(a) in which only two line segments can be discriminated. Furthermore, Fig. 10(a), (b), (c) and (d), (e), (f) show the detection results of applying our method and *LSD* to underwater raster-scan images in Fig. 1 (right column), respectively, although it is difficult to tell if there exists objects in Fig. 10(d), one can perceive two humans

Table 1 Performances of different contour depiction methods.

	Our method	Papari and Petkov (2011)	LSD (2010)	Grigorescu et al. (2004)	Canny
Pratt's (FOM)	0.54	0.39	0.44	0.33	0.13

Fig. 11. Pictorial comparison results. Left to right columns: (a) Input images. (b) Ground truth of the input images. (c) Results of Papari et al.'s method. (d) Results of the proposed method. (e) Results of *LSD*.

by looking at Fig. 10(a). Apparently, these experimental results demonstrate superiority of our method over *LSD* in depicting contours conformal to our visual perception, In addition, they justify that the capability of correctly perceiving an object is indeed mainly attributed to the well depicted curved contours.

As in (Grigorescu et al., 2004), we next use forty images comprising complex backgrounds (e.g., rivers, rocks and bushes) and objects (e.g., animals and vehicles) as input images to test various contour detectors. The Pratt's FOM (figure of merit) is adopted to obtain a matching degree between the ground truth and the detection result (Papari and Petkov, 2011). Table 1 shows that our work has fairly superiority over other methods in detecting perceptual contours. Note that the ratio value in Table 1 is acquired by averaging the corresponding scores over 40.

Moreover, as shown in Fig. 11, three nature images (first column) chosen from the forty images and the corresponding ground truth (second column) are demonstrated to show that our method can depict contours (fourth column) more consistent with the contours perceived by human visual system than other methods (third column: Results of Papari et al.'s method, fifth column: Results of *LSD*).

Furthermore, synthesized binary images in Fig. 12 (first row) are tested to demonstrate that the proposed method has the capability of depicting recognizable but not necessarily existing objects. In the synthesised binary images, V^O is normalized to [0, 1] and thresholded by 0.5, and $G^L_{Can}(x, y)$ has only two grey values, 0 and 1, resulting in the prior probability of a pixel being 1. To perform the Bayesian calculations, each $G^L_{Can}(x, y)$ is multiplied by 0.5. This assumption of an arbitrary pixel being equally likely to be or not to be a perceptual pixel is reasonable, just as the symbols 0 and 1 are equally likely to be encountered on the Internet. The results in Fig. 12 (second row) clearly show that pixels with a constant orientation are accurately detected by our method, with all the circular contours with varying orientations being ignored. In contrast, other methods cannot provide the same performance (as shown in the third and fourth rows). These results justify that the proposed method can not only effortlessly construct Gestalt objects, but also perceive salient or aligned edges in accordance with the human visual tendency.

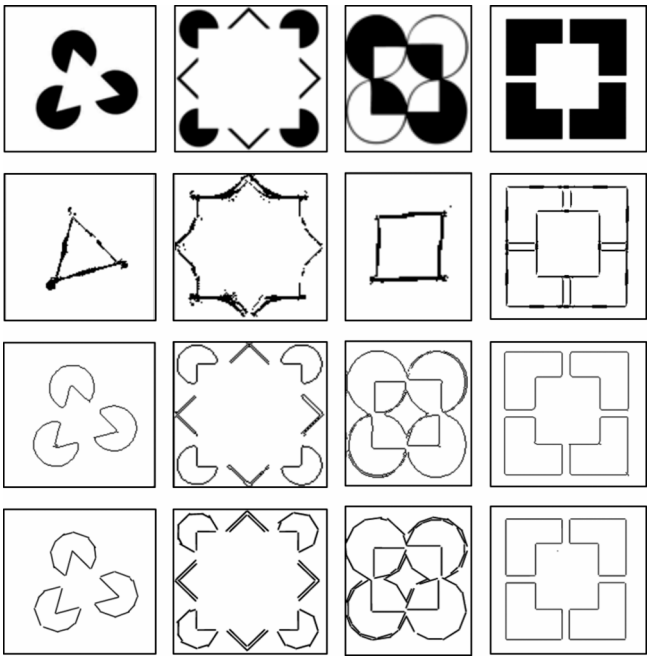


Fig. 12. Performance of the proposed method on synthesized images. From top to bottom: (First row) synthesized images. (Second row) results of applying the proposed method to the synthesized images. (Third row) results of applying the Papari et al.'s method. (Fourth row) results of applying *LSD*.

Sensitivity to parameters of σ and r is shown in Fig. 13, where each dot represents an average result of testing 10 different images randomly chosen from the aforementioned 40 images. As we can see, incrementing the value of σ only incur slight negative effect on FOM. On the other hand, although the highest FOM is obtained when $r = 5$ in Fig. 13 (b), it is not worthwhile to gain 8% in FOM performance (FOM = 0.5 when $r = 4$ and FOM = 0.54 when $r = 5$) at the expense of 58% increase in computation time (the total number of pixels within a sampling mask is 50 when $r = 4$, whereas it is 79 when $r = 5$). Therefore, r is set to 4 throughout this work.

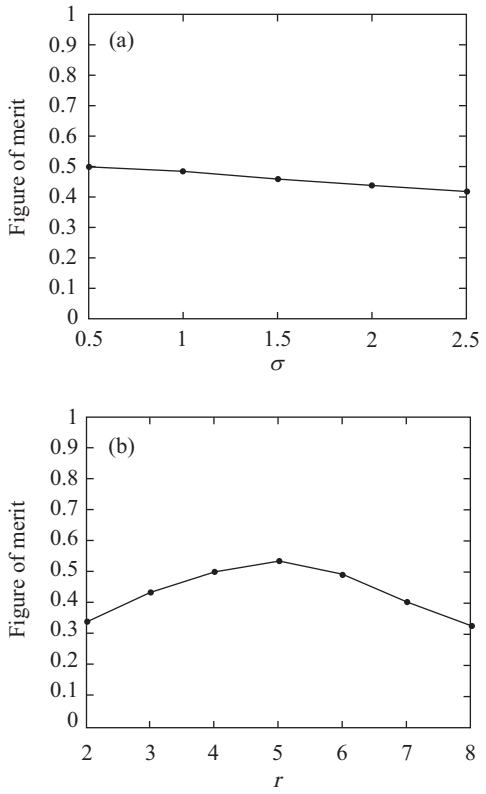


Fig. 13 (a) σ versus FOM. (b) r versus FOM.

VI. CONCLUDING REMARKS

The beauty of this work is threefold: (1) by combining *MLE* and the entropy-driven scheme, a sampling mask that optimally samples neighboring pixels and best fits the target pixel can be obtained. Characterized by taking the possible perception contradiction effect into account, the resulting directivity is useful for facilitating the decision of the salient alignment locations to be conveniently made under the Bayesian framework. (2) With a small set of parameters, one can fulfill the *Gestalt* laws of proximity and similarity simultaneously. (3) Noise and outliers can be suppressed as they have relatively small directivity.

Comparison results of extensive experiments including underwater images have justified that our method has superiority in depicting both straight and curved contours perceived by the human visual system. In particular, analysis in Section IV and experimental results in Section V have verified that the optimal directivity of individual pixel is effective in depicting perceptual contours via suppressing noise and outliers with scattering gradient magnitude and low orientation similarity, thus not only alleviating the problem that plagues aforementioned region-based perceptual contour detectors (i.e., unable to detect curves and junctions due to lacking the directivity of individual pixel relative to its neighboring pixels), but also breaking limitations of conventional edge detectors.

The sampling mask in this work has a fixed area, extensive experiments have indicated that a mask with $r > 5$ is too large to work properly for most input images. Thus, one possible future research may be directed to the determination of a dynamic mask size for improving the computation efficiency. Even though Canny detector is widely known as optimal in terms of various criteria, test results of Table 1 indicates that in some cases the Canny detector may not good enough for selecting pixels of interest.

Furthermore, as shown in Fig. 12 our method can better depict most perceptual objects than other methods, yet due to its pixel based nature, there still exists a great improvement room for the computation efficiency. Our method requires 0.016 s for processing a pixel on an Intel Core 2 Duo CPU 2.40 GHz, compared to < 0.0006 s in other methods. Also, there exists a need of a better selection strategy than the Canny detector to further improve the FOM performance of our method. Last but not least, psychology factors such as colors can be included as an additional feature for further improving the performance in the future.

ACKNOWLEDGMENTS

This work was supported by the National Science Council of Taiwan under grant NSC 102-2221-E-019-056.

REFERENCES

Azriel, R. and M. Thurston (1971). Edge and curve detection for visual scene analysis. IEEE Transactions on Computers 100, 562-569.

Bazeille, S., I. Quidu, L. Jaulin and J. P. Malkasse (2006). Automatic underwater image pre-processing. Proceedings of the SEA TECH WEEK Caractérisation du Milieu Marin.

Cai, L., L. He, T. Yamashita, Y. Xu, Y. Zhao and X. Yang (2011). Robust contour tracking by combining region and boundary information. IEEE Transactions on Circuits and Systems for Video Technology 21, 1784-1794.

Cai, J. and S. Miklavcic (2013). Surface fitting for individual image thresholding and beyond. IET Image Processing 7, 596-605

Canny, J. (1986). A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 679-698.

Desolneux, A., L. Moisan and J. Morel (2007). From gestalt theory to image analysis: a probabilistic approach. Springer-Verlag, New York.

Elder, J. H. and R. M. Goldberg (2002). Ecological statistics of gestalt laws for the perceptual organization of contours. Journal of Vision 4, article 5.

Ellis, W. D. (1999). A sourcebook of Gestalt psychology. Humanities, New York.

Foresti, G., L. (2001). Visual inspection of sea bottom structures by an autonomous underwater vehicle. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 31, 691-705.

Galceran, E., R. Campos, N. Palomeras, D. Ribas, M. Carreras and P. Ridao (2014). Coverage path planning with real-time replanning and surface reconstruction for inspection of three-dimensional underwater structures using autonomous underwater vehicles. Journal of Field Robotics 32(7), 952-983.

Gonzalez, R. C. and R. E. Woods (2007). Digital image processing, 3rd Edition. Prentice Hall, New Jersey.

Goldstein, E. (2013). Sensation and perception, 9th Edition. Cengage Learning, Kentucky.

Grigorescu, C., N. Petkov and M. A. Westenberg (2004). Contour and boundary detection improved by surround suppression of texture edges. Image and

Vision Computing 22, 609-622.
Guy, G., and G. Medioni (1992). Perceptual grouping using global saliency-enhancing operators. Proceedings of IEEE 11th International Conference on Pattern Recognition, Conference A: Computer Vision and Applications 1, The Hague, Netherlands, 99-103.
Kim, A. and R. Eustice (2009). Pose-graph visual SLAM with geometric model selection for autonomous underwater ship hull inspection. IEEE/RSJ International Conference on Intelligent Robots and Systems, Missouri, USA, 1559-1565.
King, D. B. and M. Wertheimer (2005). Max Wertheimer and Gestalt Theory. Transaction Publishers, New Jersey.
Kunz, C. and H. Singh (2013). Map building fusing acoustic and visual information using autonomous underwater vehicles. Journal of Field Robotics 30, 763-783.
Laplace, P. S. (1814). Théorie analytique des probabilités (Analytic theory of probabilities), 2nd Edition. Courcier, Paris. (in French)
Malik, J. (2006). Visual grouping and object recognition. 11th International Conference on IEEE Image Analysis and Processing, 612-621.
Mallios, A. (2014). Sonar scan matching for simultaneous localization and mapping in confined underwater environments. Ph. D. Thesis, Department of Architecture and Construction Engineering, University of Girona, Spain.
McAndrew, A., J. H. Wang and C. S. Tseng (2010). Introduction to Digital Image Processing with MATLAB, Asia Edition. Cengage Learning, Taipei.
Mittleman, J. and L. Swan (1993). Underwater inspection for welding and overhaul. Naval Engineers Journal 105, 37-42
Negahdaripour, S. and P. Firoozfam (2006). An ROV stereovision system for ship hull inspection. IEEE Journal of Oceanic Engineering 31, 551-546.
Newson, A., A. Almansa, Y. Gousseau and P. Pérez (2014). Robust automatic line scratch detection in films. IEEE Transactions on Image Processing 23, 1240-1254.
Papari, G. and N. Petkov (2011). An improved model for surround suppression by steerable filters and multilevel inhibition with application to contour detection. Pattern Recognition 44, 1999-2007.
Quidu, I., L. Jaulin, A. Bertholom and Y. Dupas (2012). Robust multi-target tracking in forward-looking sonar image sequences using navigational data. IEEE Journal of Oceanic Engineering 37, 417-430.
Ren, X., C. C. Fowlkes and J. Malik (2008). Learning probabilistic models for contour completion in natural images. International Journal of Computer Vision 77, 47-63.
Sajda, P., E. Pohlmeyer, J. Wang, L. C. Parra, C. Christoforou, J. Dmochowski, B. Hanna, C. Bahlmann, M. K. Singh and S. F. Chang (2010). In a blink of an eye and a switch of a transistor: cortically coupled computer vision. Proceedings of the IEEE 98, 462-478.
Sánchez-Nielsen, E. and M. Hernández-Tejera (2011). Real-time tracking using A* heuristic search and template updating. IET Computer Vision 5, 169-177.
Shotton, J., A. Blake and R. Cipolla (2008). Multiscale categorical object recognition using contour fragments. IEEE Transactions on Pattern Analysis and Machine Intelligence 30, 1270-1281
Stojanovic, M. (1996). Recent advances in high-speed underwater acoustic communications. IEEE Journal of Oceanic Engineering 21, 125-136.
Timotius, I. K. and I. Setyawan (2014). Using edge orientation histograms in face-based gender classification. IEEE International Conference on Information Technology Systems and Innovation (ICITSI), 93-98.
Trubuil, J., T. Le Gall, G. Lapierre and J. Labat (2001). Development of a real-time high data rate acoustic link. Proceeding of IEEE Oceans Conference and Exhibition 4, Escondido, USA, 2159-2164.
Tseng, C. S., C. T. Lin, C. W. Lin and J. H. Wang (2012). Perceptual edges preservation conformal to human vision perception. IEEE 13th International Conference on Information Reuse and Integration, Las Vegas, USA, 173-178.
Ullman, S. (1976). Filling-in the gaps: the shape of subjective contours and a model for their generation. Biological Cybernetics 25, 1-6.
Von Gioi, R. G., J. Jakubowicz, J. M. Morel and G. Randall (2010). LSD: A Fast line segment detector with a false detection control. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 722-732.
Yufit, G. and E. P. Maillard (2013). 3D forward looking sonar technology for surface ships and AUV: Example of design and bathymetry application. IEEE International Underwater Technology Symposium, Tokyo, Japan, 1-5.