Visual feature based registration of images from different retinal eccentricities

Yitzhak Yitzhaky and Eli Peli
The Schepens Eye Research Institute, 20 Staniford, Boston, MA 02114, USA.

ABSTRACT
The development of variable resolution displays that correspond to the spatial resolution of the visual system motivates the examination of basic image processing methods such as image registration, using images with different resolutions. Such examination is further motivated by the visual stability that is the perception of a stable visual world in spite of saccadic eye movements, which suggests a registration process of displaced retinal images from different eccentricities, done by the visual system. In this paper, the registration process is explored using two techniques. In one, it is done between images with different spatial resolutions as would appear in different eccentricities. In the other, the vision model is extended to produce feature-detected images from different eccentricities, which are then registered. For this purpose, a vision model-based feature extraction technique is extended to include a dependency on retinal eccentricity. The registration results show that the displacements between the varying resolution images can be identified from both gray-scale and edge-detected images. As expected, higher sensitivity to small spatial shifts appears when the edge-detected images are used.

OCIS codes: 100.2000, 100.5010, 330.2210, 330.4060, 330.6110, 330.6130
1. INTRODUCTION

Processing of images with resolution that varies according to the spatial properties of the human visual system is motivated mainly by two disciplines. One discipline is the development of displays with variable resolution similar to the resolution of the visual system that strongly declines away from the direction of gaze (foveated imaging). Such displays are most effective when the direction of gaze is tracked so that the highest resolution region at the display can be kept aligned with the highest resolution region of the eye (fovea). Implementation of such displays can decrease the bandwidth of the transmitted images and can be exploited for image compression.

The second discipline is the possible use of registration of visual information in the brain to create the appearance of a stable and aligned visual world across abrupt eye movements (saccades) that form sequences of displaced retinal images of the scene. This phenomenon referred to in the literature as visual stability or space constancy was widely explored experimentally.

Image registration in general is the process that determines the best match between different acquired images of the same scene. Automatic image registration is required in many image processing applications when the “same” scene is acquired with different sensors, at different times, from different viewpoints, and so on. Extensive study of image registration in various fields of image processing has been carried out including several review articles.

In this work we examine the use of variable resolution visual information as a basis for an image registration process. Registration of visual information can verify the viability of the use of such
information as a basis for visual stability, and can be employed when visually varying resolution systems are considered.

Two types of visual information are developed here, and employed in the spatial variant registration process: one is a pair of arbitrarily shifted images with different spatial resolutions as if obtained from different retinal eccentricities, and the other is a pair of feature (edge-detected) images, obtained from these shifted images, according to a multi channel visual model based feature extraction technique from early cortical processing. To simulate the visual information at different retinal eccentricities, we extend the space invariant feature extraction method to include the spatially variant properties of the visual system, and an estimation of actual image quantities that transfer through each visual channel at different eccentricities. These quantities were used to form the images at different eccentricities, and as a basis to decide what channels can be used in the edge detection process at different eccentricities. Registration results in which the images from different eccentricities are used, are compared to registration results in which the feature images are used, since both may be available in the visual system at different levels.

In the next section the use of retinal information as a basis for visual stability is addressed. In Section 2 the visual model-based feature extraction method is summarized. An extension of the method, that incorporates the space variant properties of the visual system and the properties of the actual observed image, is presented in Section 3. Feature extraction and registration of images from different eccentricities are presented in Section 4, including a basic digital image registration process, registration results, and a quantitative registration performance measure. Conclusions are presented in Section 5.
1.1 The use of visual information as a basis for visual stability

Several types of information sources have been proposed to enable the visual stability. These information sources can be divided into two categories: retinal and extra-retinal. Extra-retinal signals include a copy of neural outflow from the brain to the extra-ocular muscles (an efference copy), which encodes the initiation and intended extent of the eye movement, and neural inflow from stretch receptors in the extra-ocular muscles (proprioceptive signals), which encodes the occurred eye movement. Retinal information encodes the movement of the retinal image.

Various theories have been suggested to explain the visual stability perception across saccades. From research and reviews that examined the visual stability phenomenon, it appears that any one of the information sources by itself cannot account for this phenomenon, rather both retinal and extra-retinal information (efference copy) are probably employed to achieve stability perception.

The information that is probably required for visual stability is the exact extent of the eye movement. Although such information is contained in the extra-retinal signals, its precision is too low to account for research results on visual stability. Thus, the higher resolution retinal information may be employed to refine the registration. Thus the extra-retinal information might only restricts or limits the range over which the search for exact registration is needed.

As a result of the high speed of movement during saccades, a saccadic eye movement can be considered as producing two separate successive static retinal images, a new one from the current fixation (post-saccadic image) and a reminder from the previous fixation (pre-saccadic image). Heiner et. al concluded that visual stability depends on comparison of common elements in the
pre- and post-saccadic images. Bahcall and Kowler concluded that signals representing intended saccades initiate a visual comparison of pre- and post-saccadic images used to maintain perceptual stability across saccades and to generate an oculomotor error signals that ensure fixation accuracy.

The pre- and post-saccadic images are shifted, one with respect to the other, according to the eye movement extent. This means that at each point of the scene (object) space, each of the two images is represented at different spatial resolution with respect to the other as a result of the different visual spatial resolutions at different eccentricities. For example, an image feature in the space that previously appeared at the fovea, will appear subsequently at a higher eccentricity determined by the eye movement extent, and thus, be represented at a lower visual resolution.

This means that images with different spatial resolutions have to be evaluated by the visual system in order to derive the required registration information. A lower resolution image may limit the accuracy of the information that can be extracted from it. However, the spatial frequency multi-channel structure of the visual system enables an extraction of relatively high-resolution components in the image (object boundaries) that can be used for the identification of the shift between the images. The shift between the images that represents the eye movement extent is extracted in this work from the pre- and post-saccadic visual information via an image registration process.
2. A VISUAL MODEL BASED FEATURE EXTRACTION

The feature extraction technique used here is based on contemporary vision models of visual information processing. A detailed description of this feature extraction technique applies as spatially invariant is presented in elsewhere. This section presents a brief summary of this technique. A spatially variant extension is provided in the next section.

The image is first filtered through a multi-scale set of visual-channels like band pass filters. The filters are one-octave wide and separated by one-octave in their center frequency. The $i^{th}$ order filter $C_i$ (Fig. 1) used to generate the $i^{th}$ scale band pass filtered image $F_i$ applied in the frequency domain is:

$$C_i(f_x, f_y) = \begin{cases} 
0.5[1 + \cos(\pi \log_2 r - \pi)], & 2^{i-1} \leq r \leq 2^{i+1} \\
0, & \text{elsewhere}
\end{cases}, \quad (1)$$

where $f_x$ and $f_y$ are the horizontal and vertical spatial frequency coordinates, and $r$ is the radial spatial frequency ($r = \sqrt{f_x^2 + f_y^2}$). The band pass filtered images $F_i(f_x, f_y)$ are then converted to the space domain $F_i(x, y)$ and thresholded according to the contrast threshold function of the visual system. At each scale $i$, the threshold value $Th_i$ at its spatial center frequency is determined and applied as follows

$$T_i(x, y) = \begin{cases} 
+1, & \text{if } F_i(x, y) \geq +Th_i \\
0, & \text{if } -Th_i < F_i(x, y) < +Th_i \\
-1, & \text{if } F_i(x, y) \leq -Th_i
\end{cases}, \quad (2)$$
producing a tri-level thresholded $i^{th}$ scale $T_i(x, y)$ at each scale. A second tri-level image $E_i$ of detected visual features is then obtained based on the correspondence across several scales. This stage represents an all or none decision process carried out using the visual information arriving from the previous stages.

$$E_i(x, y) = \begin{cases} 
+1, & \text{if } T_i(x, y) = +1, \forall i \\
0, & \text{otherwise} \\
-1, & \text{if } T_i(x, y) = -1, \forall i 
\end{cases} \quad (3)$$

The resulting detection differs from conventional edge detection techniques in that the edge features are represented in pairs, with a black pixel located on the darker side of the edge and a white pixel located on its brighter side (for example, see Figs. 5 and 8).

3. AN EXTENTION TO WIDE VISUAL FIELD

The above model is suitable for visual processing within a small region of the visual field where visual functioning is reasonably homogeneous. Yet, space variance should be taken into account when it comes to modeling visual processing within larger visual field areas. Empirical research of the visual system suggests a multi-channel vision model, with channels varying according to the spatial location (mainly eccentricity) with respect to the fovea$^{17-18}$ and in which several channels exist at each position in the visual field$^{19}$. Regarding different variations that have been reported about the manner of the change of the receptive channels with eccentricity, it was shown that the change of the contrast sensitivity function (CSF) with eccentricity differs for different
Peli et al. developed a mathematical model for the change of the contrast threshold according to the spatial frequency and the eccentricity from the fovea. Two properties of the visual system responding to changes in frequency and eccentricity dominate the model. The first is the exponential drop with eccentricity of measured contrast thresholds at any one spatial frequency. The second, called contrast constancy, is the invariance of the appearance of (suprathreshold) objects with changes in size of their retinal images, which results from their distance from the eye. The inverse relation between the retinal size and the spatial frequency, while contrast constancy is maintained, suggests that contrast thresholds vary as a product of the spatial frequency and the eccentricity. The contrast threshold $Th$ at eccentricity $\theta$ and frequency $r$ is expressed by this model as

$$\ln[Th(\theta, r)] = a\theta r + \ln[Th(0, r)],$$

where $a$ is a constant and $Th(0, r)$ is the foveal threshold at frequency $r$. This model was successfully fitted to results of contrast thresholds measurements at different frequencies as a function of eccentricity, reported by various studies, and was recently verified empirically. This model can be applied to derive the cortical image obtained from different retinal eccentricities.

By incorporating the different contrast thresholds at different eccentricities into the multi-channel visual model, one can evaluate the spatial signal quantity that can be received by the eye at every channel $i$ and eccentricity $\theta$, according to the area captured between the filter and a contrast threshold. This area is an extension of the definition of the Modulation Transfer Function Area (MTFA), which is the area between the frequency response of an optical system and the
contrast threshold of the visual system. The MTFA gives a quantitative single number measure of the possible quality of an image transferred through a system and shown to the human eye. Hence, we define a single Channel MTFA (CMTFA) as:

\[
CMTFA_{\theta,j} = \int_{2^{j-1}}^{2^j} [C_i(r) - Th(\theta, r)] dr, \quad C_i(r) > Th(\theta, r). \quad (5)
\]

A one-dimensional intersection of the incorporation of the contrast thresholds at different eccentricities with the multi-channel visual model is presented in Fig. 1. The lower dotted line is the contrast threshold at the fovea, the continuous lines are contrast thresholds at eccentricities from 5 degrees (the second lower) to 30 degrees (the upper), and the dashed lines are the receptive channels (filters) according to equation (1). The \( CMTFA_{\theta,j} \) is expressed in this figure as the area captured between the spatial filter \( i \) and the contrast threshold of eccentricity \( \theta \). For example, \( CMTFA_{4,2} \) is marked in Fig. 1 by vertical lines. Table 1 shows values of \( CMTFA_{\theta,j} \) for channel orders and eccentricities shown in Fig. 1.

The CMTFA values in Table 1 reflect known properties of the visual system. The CMTFA is reduced with eccentricity because the contrast threshold is higher as eccentricity increases. The CMTFA is increased with regard to the channel order since the bandwidths of the channels increases logarithmically as the order increases. However, the increase of the CMTFA with the order is smaller than the logarithmic increase of the filter size since the contrast threshold also increases with the channel order. These CMTFA values can be explained by the properties of the visual system that evolved with regard to natural scenes.29 Though the important and useful
information in the image is mainly the boundaries of the objects, the areas of the boundaries are obviously much smaller than the homogenous parts in the image. Thus, the boundaries emit a much smaller energy. Therefore, it is the interest of the visual system to enable wider receiving bandwidths for the higher frequency channels that receive the boundary (high frequency) information. This is true at the lower eccentricities, mainly in the fovea. The decrease of the CMTFA with eccentricity escalates at the higher channel orders since these orders do not dominate the higher eccentricities where accurate information is less important. At higher eccentricities (used mainly for navigation and danger attention), a rough knowledge of objects existence (available by their lower frequencies) is sufficient, and therefore, no image energy receiving capability (represented by the CMTFA values) is required there at the higher frequencies. This setup of the visual system is necessary to efficiently manage the huge amount of optical information constantly reaching the visual system according to natural needs.

Our purpose here is not just to realize the visual channels performance capabilities according to eccentricity, represented by the CMTFA, but also to realize the actual image quantities transferred by each of the visual channels toward later visual processing stages. These image quantities can help us determine which visual channels may be used as an information source for registration of images from different eccentricities. Inclusion of an input real image signal spectrum in the visual receiving setup with different eccentricities is shown figuratively in one dimension in Fig. 2. The frequency range is based on an image of 256 x 256 pixels spanning a visual field of 4°. The spectrum shown is the radial average of the 2-D spectrum of the image shown in Fig. 4a, and it resembles conventional models of the radial average spectrum of real natural images known to be decreasing functions of the spatial frequency, usually proportional to 1/r.
Since the 1-D image spectrum plotted in Fig. 27 is the average of both different spatial locations and different spatial frequency orientations, it does not show the actual amount of image intensity transferred by each channel. For instance, although the average spectrum at the high frequencies is below the threshold, some spatial locations with strong high frequency content (sharp high contrast edges) may elicit a response by the visual system’s higher order channels.

The contrast threshold value $Th_i$ used in the spatial domain in equation (2) can now be generalized to different eccentricities by the value $Th_{\theta,i}$, thus, equation 2 becomes:

$$T_{\theta,i}(x,y) = \begin{cases} +1, & \text{if } F_i(x,y) \geq +Th_{\theta,i} \\ 0, & \text{if } -Th_{\theta,i} < F_i(x,y) < +Th_{\theta,i} \\ -1, & \text{if } F_i(x,y) \leq -Th_{\theta,i} \end{cases}$$

(6)

where

$$Th_{\theta,i} = 2^{\left\lfloor a \theta 2^i + \log_2(Th_{\theta,j}) \right\rfloor}$$

(7)

according to equation (4) with $r = 2^i$. $Th_{\theta,i}$ is an approximation (median value) of $Th(\theta, r)$ that changes throughout the frequency band of the channel, from $2^{i-1}$ to $2^{i+1}$. At every channel, the filtered image intensity that is above the contrast threshold, $T_{\theta,i}(x,y)$, is integrated, and the result represents the amount of intensity from the image, passed through this channel for later visual processing. This information is used here to determine which channels should be employed in the feature detection process forming the feature space for the registration process. Table 2 presents
the energies of the image of Fig. 4a, which passed through each channel and different eccentricities.

It can be seen from Table 2 that although the CMTFAs of the higher order channels are bigger (Table 1), the levels of image intensities that pass through these channels are smaller. The reason is that in natural scenes, high frequencies (roughly object borders) occupy smaller areas in the image, and usually have smaller intensities. Thus, they have less capability to exceed the increased contrast thresholds at the higher frequencies.

**Feature extraction implementation for different eccentricities**

The visual model based feature extraction technique presented in Section 2 employs several filters that correspond to different visual channels. The actual filters used depend on the frequency distribution of the scene and on the frequency response of the visual system, which vary according to the eccentricity. The spatial angular frequency range of a digital image depends on the size of the image (in pixels) and its angular span. Only channels that transfer a significant signal quantity are employed for the feature detection purpose. Combined information from several channels is used for feature detection according to equation (3). Channels with zero intensity in Table 2 obviously cannot be used for the feature detection task. The minimal intensity value that allows the use of a channel can be found out experimentally. Since higher orders provide the finer objects details, it is preferable to use as higher channel order as possible, limited by the requirement of significant intensity that passes that channel. Lower orders are required mostly to reduce the noisiness of the highest order information and have small contribution for the feature detection. The resulting edge detection (equation 3) becomes:
\[
E_{\theta,i}(x,y) = \begin{cases} 
+1, & \text{if } T_{\theta,i}(x,y) = +1, \forall i \in \mathcal{R} \\
0, & \text{otherwise} \\
-1, & \text{if } T_{\theta,i}(x,y) = -1, \forall i \in \mathcal{R} 
\end{cases}
\]

where \(\mathcal{R}\) is the set of channels used according to Table 2. The number of channels in the fovea that we included in \(\mathcal{R}\) was four.\(^{17,10}\) This number may be reduced with the increase in eccentricity because fewer channels are available there (Tables 2) and the lowest channels have no role in the detection of edges that correspond to the higher frequencies in the image.

4. REGISTRATION OF IMAGES FROM DIFFERENT ECCENTRICITIES

With one saccadic transaction, the fovea at first is directed and fixated at a certain location. At this time, different locations on the retina receive image information from different directions of the wide field of view of the eye forming a perceived scene with space variant resolution according to eccentricity. After an abrupt saccadic movement of the eye toward a different fixation location, different areas with different resolutions are “placed” one upon each other across contiguous saccades; yet, the perceived image is continuous and stable. An illustration of pre- and post-saccadic images is presented in Fig. 3. A region of the visual field where the man’s head is at the fovea is shown in Fig. 3a. Fig. 3b shows a displaced region following a movement of the fovea from the man’s head to the woman’s head. Identification of the eye movement extent may be carried out by comparing these two images in a registration process. A square displaced image segment (of the woman’s head) with low spatial resolution from the image in 3a is shown at the right side of the image in 3b. This segment represents an area from one image that can be
compared to areas with different resolutions in the second image, in an image registration process.

In this section we perform and compare registration, using segments from different eccentricities using two types of image information. One type is pre- and post-saccadic visual images, and the second type, is edge-detected features extracted from the same images according to the visual model presented previously. The processed images registered, were simplified by assuming that they contain locally space invariant resolutions, which imply that the image segments used in the registration process span small visual field.

4.1. Digital image registration

In order to put the visual information based registration performed here in a conventional context of image registration, we will briefly review the registration process using four basic components: the feature space which is the information from the images used to perform the matching between the images, the search space that is the potential transformations that establish the correspondence between the images (for example shift, rotation, rescaling), the search strategy that decides how to choose the next step, and the similarity metric which determines the match measure between one image and the other transformed image (for example correlation, mean squares, absolute differences). In this work, visual information will be used as a feature space; the search space will be all the possible shifts between the images within an assumed maximum displacement that may be defined according to the extra-retinal information; the search strategy will simply be scanning all possible shifts; and the similarity metric will be the sum of absolute values of differences (SAVD). In the image registration process a subsection (template) in one image is compared to shifted subsections with the same size throughout a search area in the second image, according to the
similarity measure. The shift that gives the best match is the registration result. Assuming a maximum possible displacement size of $L$ pixels between the images, non-overlapping areas between the images may exist in a frame with an $L$ width, at the edges of the images. The search area size should be larger than the template size so that the template can be shifted within it by at least $L$ pixels in all directions in order to cover all the possible displacements between the images. Thus, the minimum search area length is $2 \cdot L + \text{template length}$, where the template size should be large enough to include sufficient spatial salient features for registration (non-homogenous areas). Given a search area of size $J \times K$ and enabling assumption of movement in all directions, the normalized SAVD between a template $A$ of size $J \times K$ and a shifted subsection $B$ of the same size in the second image is:\textsuperscript{31}

\[
SAVD(m,n) = \sum_{j=-J/2}^{J/2-1} \sum_{k=-K/2}^{K/2-1} \left( A_{jk} - \overline{A}_{jk} \right) - \left( B_{j-m,k-n} - \overline{B}_{j-m,k-n} \right) \tag{9}
\]

where $m$ and $n$ are the horizontal and vertical shifts in the search areas, and $\overline{A}_{jk}$ and $\overline{B}_{j-m,k-n}$ are the averages of the gray levels around $A_{jk}$ and $B_{j-m,k-n}$ respectively. The subtraction of the averages normalizes the subsections to have zero average, and improves the reliability of the registration since it decreases the effect of different average luminance levels between the images. The registration point is the values of $m$ and $n$ where the SAVD has a minimum value.

4.2 Simulation of image registration from different eccentricities

A first image was taken by a digital camera, and was simulated to appear as recorded at the fovea ($\theta = 0$) with $4^\circ$ visual span.\textsuperscript{21} Another image from the same scene taken later at an unknown shift
from the first image was used as the misaligned image resulting from a saccadic eye movement. This image was simulated as if recorded at a retinal eccentricity $\theta = 10^\circ$. Both images represent locally space invariant segments of the visual field with relatively small visual span. The image at the fovea is shown in Fig. 4a, and the image at eccentricity $10^\circ$ is shown in Fig. 4b. These images have been produced by summing the thresholded (non-linearly filtered) image scales from all the channels at eccentricities $0^\circ$ and $10^\circ$ (the quantities of these scales appear in Table 2). Edge detection results, obtained by implementing the vision model based feature extraction technique (equation 8) at the different eccentricities, are presented in Fig. 5. Fig. 5a is an implementation at the fovea where the center frequencies of the channels employed were 2, 4, 8 and 16 cycles/deg. In Fig. 5b for $10^\circ$ eccentricity, the center frequencies of the channels employed were 1, 2, 4 and 8 cycles/deg. The selected channels are marked in Table 2 by the gray frames. The channels that give the most usable accurate feature information are the highest channels with significant intensity. Fig. 6a shows the result of the registration process using the intensity images shown in Fig. 4, and Fig. 6b shows the registration result using the edge detections shown in Fig. 5. The registration maps in Fig. 6 are a normalized SAVD (equation 9) at every shift distance within a 60 x 60 pixels assumed maximum possible displacement between the images ($\pm 30$ pixels in each horizontal and vertical directions). A darker pixel represents better similarity between the matched features. It can be seen that the registration map of the edge-detected features has a significantly sharper peak at the registration point, but it is less smooth than the registration map using the intensity images. The sharpness property of the registration point relative to its surrounding is expected when thin features such as object contours are used as a feature space, however, the bi-polar nature of the visual model based detected features used here, usually further increase the relative sharpness of the registration point.
Another example is shown in Figs. 7-9. In this case the registration is between foveal and 20° eccentricity simulations of images (Figs. 7a and b respectively) taken with unknown shift between them. Feature extraction according to the visual model is presented in Figs 8a and b for foveal and eccentric images respectively. The channels used for feature extraction in this case were the same as in the previous case for the foveal image, and for 20° eccentricity the center frequencies of the channels were 2, 4 and 8 cycles/deg. These channels are not selected according to table 2 that shows the energies of the image of Fig. 4 and not the images of Fig. 7. The registration match maps are presented in Figs. 9a and 9b for the cases of using the intensity image and the edge detection, respectively. In this case too, a sharper registration peak can be seen in Fig. 9b where the edge detection was used for registration.

4.3. Registration performance measure

The visual stability of the visual system implies a possible quick and accurate registration process. The differences between the appearances of the registration maps that result using edge detection images (Figs. 6b and 9b) versus intensity images (Figs. 6a and 9a) may imply differences in possible performances of the visual system. A sharper registration point (a smaller darker area relative to its environment) suggests higher sensitivity to small spatial shifts. A quantitative measure of this sharpness quality can be obtained by fitting a parametric Gaussian function to the registration map. When a fitting is performed in the vertical and the horizontal direction, a measure of the sharpness of the registration point can be the average standard deviation from both directions. Fig. 10 presents Gaussian fit measures of the registration results in Fig. 6, showing the sharpness of the registration point relative to its neighborhood. Fig. 10a is
the fit for the registration result using the intensity images (Fig. 6a), producing an average standard deviation 10.2. Fig. 10b is the fit for the registration result using the edge detections (Fig. 6b), producing an average standard deviation 4.5. The Gaussian fits of the second example are shown in Fig. 11. The average standard deviations in this case are 10.7 where the intensity images were used, and 3.8 where the edge detections were used.

5. SUMMARY AND DISCUSSION

In this paper, a visual based model for feature extraction was extended to include the different properties of the visual system at varying retinal eccentricities. Properties of a real image were incorporated showing quantitatively the combination of wider channels at higher spatial frequencies, which receive lower image intensity inputs obstructed by bigger contrast thresholds at higher eccentricities and spatial frequencies. Such model extension enabled the performance of simulations of processes that involve information from different eccentricities.

Image registration from different eccentricities implicitly performed by the visual system across saccades was simulated using visual information according to the extended model. The shifted images from different eccentricities, and the edge-associated features extracted from these images according to the model, were separately used as inputs in the registration process. The extent of the misalignment between the images from different eccentricities was identified when either the images, or the extracted features, were used. Comparing the use of the two different inputs, the identified misalignment extents were usually identical, or with one pixel difference. Such results can be considered to be similar since the actual misalignment extent is most likely not an integer number of pixels and small differences between the misalignment extents occur across the image, since the imaging process is not perfectly space invariant. In another scope of the comparison
between the uses of the two input types (section 4.3), we measured the sharpness of the registration point relative to it’s surrounding in the registration map, according to the width of a Gaussian fit to the registration map. The smaller Gaussian width obtained when the extracted features were used means that in this case, the misalignment extent identification process has higher sensitivity to small spatial shifts. Therefore, it can be concluded that when the extracted edge-associated features are used, a higher certainty about the exact misalignment extent is achieved. Although the examples shown here register images from the eccentricity and the fovea, successful registration was achieved when images were simulated from variety of eccentricities up to 30 degrees.

Implementation of a vision model in an image processing application may benefit both disciplines. In this work, in the vision discipline, it shows a quantitative evaluation of the interaction between real images and the visual system, and it demonstrates a possible use of the visual information (according to a vision model), as an aid to achieve visual stability. In the image processing discipline, a new technique for registration of images with different spatial resolutions was developed, and successfully implemented using real, arbitrarily shifted images with significantly different spatial resolutions. This can be used in cases of foveated imaging systems, in which the spatial resolution is similar to the space variant resolution of the visual system.

ACKNOWLEDGMENTS

Supported in part by NIH grants EY05957 and EY12890.
REFERENCES


Table captions

1. Values of $CMTFA_{\theta,j}$ for channels and eccentricities shown in Fig. 1.

2. Integrated energies of the image of Fig. 4a, which passed through each channel at different eccentricities.

Figure captions

1. Cross-sectional profiles of contrast thresholds $Th(\theta,r)$ at different eccentricities with the multi-channel filters $C_i(r)$ of the visual model; the lower dotted line is the contrast threshold at the fovea, the continuous lines are contrast thresholds at eccentricities from $5^\circ$ (the second lower) to $30^\circ$ (the upper), and the dashed lines are the receptive channels responses according to (1). The area marked by the vertical lines is the $CMTFA_{15,2}$ (at eccentricity $15^\circ$ and the channel with 4 cycles/deg center frequency).

2. Same as Fig. 1, but with the average radial spectrum of a real input image that was used here for registration. The frequency range is based on an image of 256 x 256 pixels spanning a visual field of $4^\circ$.

3. An Illustration of pre- and post-saccadic images: (a) the man’s head is at the fovea; (b) the woman’s head is at the fovea; a square displaced segment of the woman’s head with low resolution from the upper image is shown at the right side. This segment represents an area from one image that can be compared to areas with different resolutions in the second image in the registration process.
4. Images of the same scene recorded with unknown spatial shift between them, and simulated for appearances at different retinal eccentricities of (a) 0° and (b) 10°. These images have been produced by integrating the filtered image scales from all the channels at each eccentricity.

5. An implementation of the vision model based feature extraction technique to the images of Fig. 4 at different eccentricities (equation 8); (a) at the fovea, and (b) at 10° eccentricity.

6. The registration results (normalized SSAVD values for different shifts between the images) using the different feature spaces: (a) the images shown in Fig. 4; identified registration point is at a shift (5,26), and (b) the edge detections shown in Fig. 5; identified registration point is at a shift (4,26).

7. Images of the same scene recorded with unknown spatial shift between them, and simulated for appearances at different retinal eccentricities of (a) 0° and (b) 20°.

8. An implementation of the vision model based feature extraction technique to the images of Fig. 7 at: (a) the fovea, and (b) 20° eccentricity.

9. The registration results using the different feature spaces: (a) the images shown in Fig. 7; identified registration point is at a shift (28,2), and (b) the edge detections shown in Fig. 8; identified registration point is at a shift (28,2).
10. Gaussian fits to the SAVD registration results in the registration (Fig. 6) showing the sharpness of the registration point relative to its neighborhood; (a) a fit to the result in Fig. 6a; the average standard deviation is 10.2, (b) a fit to the result in Fig. 6b; the average standard deviation is 4.5.

11. Same as Fig. 10, but for the SAVD registration results of Fig. 9; (a) a fit to the result in Fig. 9a; the average standard deviation is 10.7; (b) a fit to the result in Fig. 9b; the average standard deviation is 3.8.
Table 1. Values of $CMTF_{\theta,j}$ for channels and eccentricities shown in Fig. 1.

<table>
<thead>
<tr>
<th>Eccentricity</th>
<th>Center freq. [c/deg]</th>
<th>0.5</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>16</th>
<th>32</th>
</tr>
</thead>
<tbody>
<tr>
<td>0°</td>
<td>0.2950</td>
<td>1.192</td>
<td>4.77</td>
<td>18.9</td>
<td>73.6</td>
<td>278</td>
<td>515</td>
<td></td>
</tr>
<tr>
<td>5°</td>
<td>0.2948</td>
<td>1.189</td>
<td>4.74</td>
<td>18.5</td>
<td>63.4</td>
<td>105</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>10°</td>
<td>0.2945</td>
<td>1.186</td>
<td>4.71</td>
<td>17.6</td>
<td>40.7</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>15°</td>
<td>0.2941</td>
<td>1.182</td>
<td>4.66</td>
<td>15.9</td>
<td>18.9</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>20°</td>
<td>0.2937</td>
<td>1.177</td>
<td>4.59</td>
<td>13.3</td>
<td>3.2</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>25°</td>
<td>0.2931</td>
<td>1.171</td>
<td>4.48</td>
<td>10.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>30°</td>
<td>0.2925</td>
<td>1.163</td>
<td>4.32</td>
<td>7.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Channels are indicated by their center frequency values, and retinal eccentricities are in degrees.

All the values are multiplied by $10^3$ for a better precision representation.
Table 2. Integrated energies of the image of Fig. 4a, which passed through each channel at different eccentricities.

<table>
<thead>
<tr>
<th>Eccentricity</th>
<th>Center freq. [c/deg]</th>
<th>0.5</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>16</th>
<th>32</th>
</tr>
</thead>
<tbody>
<tr>
<td>0°</td>
<td>213</td>
<td>121.3</td>
<td>122.2</td>
<td>92</td>
<td>69</td>
<td>50</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>5°</td>
<td>213</td>
<td>121.3</td>
<td>122.2</td>
<td>92</td>
<td>62</td>
<td>5.6</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>10°</td>
<td>213</td>
<td>121.3</td>
<td>122.1</td>
<td>91</td>
<td>35</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>15°</td>
<td>213</td>
<td>121.3</td>
<td>122.0</td>
<td>90</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>20°</td>
<td>213</td>
<td>121.3</td>
<td>121.9</td>
<td>87</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>25°</td>
<td>213</td>
<td>121.3</td>
<td>121.6</td>
<td>78</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>30°</td>
<td>213</td>
<td>121.2</td>
<td>121.1</td>
<td>59</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

All values are multiplied by $10^3$ for a better precision representation. The gray frames show the set of channels $\mathcal{R}$, used in the feature extraction process for the image of Fig. 4a.
Fig. 1. Cross-sectional profiles of contrast thresholds $Th(\theta, r)$ at different eccentricities with the multi-channel filters $C_i(r)$ of the visual model; the lower dotted line is the contrast threshold at the fovea, the continuous lines are contrast thresholds at eccentricities from $5^0$ (the second lower) to $30^0$ (the upper), and the dashed lines are the receptive channels responses according to (1). The area marked by the vertical lines is the $CMTFA_{15,2}$ (at eccentricity $15^0$ and the channel with 4 cycles/deg center frequency).
Fig. 2. Same as Fig. 1, but with the average radial spectrum of a real input image that was used here for registration. The frequency range is based on an image of 256 x 256 pixels spanning a visual field of 4°.
Fig. 3. An Illustration of pre- and post-saccadic images: (a) the man’s head is at the fovea; (b) the woman’s head is at the fovea; a square displaced segment of the woman’s head with low resolution from the upper image is shown at the right side. This segment represents an area from
one image that can be compared to areas with different resolutions in the second image in the registration process.

Fig. 4. Images of the same scene recorded with unknown spatial shift between them, and simulated for appearances at different retinal eccentricities of (a) 0° and (b) 10°. These images have been produced by integrating the filtered image scales from all the channels at each eccentricity.

Fig. 5. An implementation of the vision model based feature extraction technique to the images of Fig. 4 at different eccentricities (equation 8); (a) at the fovea, and (b) at 10° eccentricity.
Fig. 6. The registration results (normalized SSAVD values for different shifts between the images) using the different feature spaces: (a) the images shown in Fig. 4; identified registration point is at a shift (5,26), and (b) the edge detections shown in Fig. 5; identified registration point is at a shift (4,26).

Fig. 7. Images of the same scene recorded with unknown spatial shift between them, and simulated for appearances at different retinal eccentricities of (a) 0° and (b) 20°.
Fig. 8. An implementation of the vision model based feature extraction technique to the images of Fig. 7 at: (a) the fovea, and (b) 20° eccentricity.

Fig. 9. The registration results using the different feature spaces: (a) the images shown in Fig. 7; identified registration point is at a shift (28,2), and (b) the edge detections shown in Fig. 8; identified registration point is at a shift (28,2).
Fig. 10. Gaussian fits to the SAVD registration results in the registration (Fig. 6) showing the sharpness of the of the registration point relative to its neighborhood; (a) a fit to the result in Fig. 6a; the average standard deviation is 10.2, (b) a fit to the result in Fig. 6b; the average standard deviation is 4.5.

Fig. 11. Same as Fig. 10, but for the SAVD registration results of Fig. 9; (a) a fit to the result in Fig. 9a; the average standard deviation is 10.7; (b) a fit to the result in Fig. 9b; the average standard deviation is 3.8.