

# Automatic Image Retargeting with Fisheye-View Warping

*Feng Liu and Michael Gleicher*  
Department of Computer Sciences  
1210 West Dayton St.  
Madison, WI 53706 USA  
{fliu|gleicher}@cs.wisc.edu

## ABSTRACT

Image retargeting is the problem of adapting images for display on devices different than originally intended. This paper presents a method for adapting large images, such as those taken with a digital camera, for a small display, such as a cellular telephone. The method uses a non-linear fisheye-view warp that emphasizes parts of an image while shrinking others. Like previous methods, fisheye-view warping uses image information, such as low-level salience and high-level object recognition to find important regions of the source image. However, unlike prior approaches, a non-linear image warping function emphasizes the important aspects of the image while retaining the surrounding context. The method has advantages in preserving information content, alerting the viewer to missing information and providing robustness.

**ACM Classification** H5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

**General Terms** Algorithm, Human Factors

**Keywords:** Image retargeting, Image warping, Fisheye-view warping, Salience map, Focus plus context

## 1 INTRODUCTION

The display of images at a small size is important for both mobile devices, such as PDAs and cellular phones, as well as for browsing image databases. Large source images, such as those taken by digital cameras, must often be adapted so that they are more effective when viewed at a small size. We call such adaptation *Image Retargeting*. Uniformly scaling images to fit the new target size provides a simple retargeting method, but such a naïve approach is flawed in that important aspects of the image may be shrunk so much they cannot be recognized, and aspect ratio changes may cause anisotropic stretching that makes important image objects unrecognizable. Recent work on image retargeting has introduced automatic methods that identify important aspects of images and create target images that emphasize these important parts. These methods rely on cropping to discard parts of the image that are deemed less important.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

UIST'05, October 23-27, 2005, Seattle, Washington, USA.  
Copyright 2005 ACM 1-59593-023-X/05/0010 ...\$5.00.

This paper introduces a new method for image retargeting based on fisheye-view warping. Like the previous cropping-based approaches, the method automatically identifies important aspects of the source image to emphasize. However, rather than completely discarding less important aspects of the image, it uses a non-linear image warping function that de-emphasizes them. The idea of using an idealized fisheye warp is a common method in information visualization to provide a “focus+context” display that both emphasizes the focus area, while retaining its context.

Our contribution is to apply the information visualization concept of a fisheye lens to the image retargeting problem. To do this, we provide methods for connecting image importance computations to fisheye warping, as well as introducing new fisheye warps that are well-suited to the retargeting problem. We also provide an empirical evaluation to show the viability of the approach.

The fisheye view-based image retargeting method offers a number of advantages over the previous cropping-based approaches. First, it provides a result that retains the context surrounding the region of interest (ROI) of the image, allowing, for example, the viewer to see not only who is in an image, but also to see where the person is. Second, the distortions alert the viewer that there is an aspect of the image that is missing or unclear, unlike cropping, where the viewer may not even know that there are aspects of the image removed. Third, fisheye warps provide improved robustness. While the method may distort an image feature mistakenly identified as unimportant, a cropping-based method would discard it altogether.

After a discussion of related work, Section 3 describes how image retargeting is automated by combining image analysis methods to determine the region of interest (ROI). This section provides details on how these methods apply to automatic cropping. Section 4 describes how warping is used for image retargeting, and introduces our fisheye techniques. We discuss an empirical evaluation of our methods in Section 5.

## 2 RELATED WORK

The problem of automatic retargeting of large images to small-size displays has been considered by a number of authors. For example, Suh et al. [19] presented an automatic thumbnail cropping algorithm. They use a variety of methods for determining the region of interest, and then crop the image to remove all parts except the ROI. Their evaluation confirms the superiority of automatic cropping over naïve

scaling. Chen et al. [4] presented an automatic retargeting method based on a richer analysis of the image. Their approach ultimately crops the image. These successes with a cropping approach motivate our method.

Setur et al. [16] present preliminary results on an automatic retargeting method for images with multiple regions of interest. Their method crops less important image regions around the periphery, and obscures less important aspects, causing a loss of context similar to cropping.

Several authors have explored interactive methods for browsing large videos and images on small devices. Fan et al. [5] use a visual attention model to find the focus region, and provide automatic, semi-automatic and manual modes for users to select and zoom into ROIs while browsing. Similarly, Liu et al. [13] use ROI information to automatically create a tour of the image. While interactive browsing may be effective in some situations, image retargeting is required when interaction is inappropriate. For example, when the viewer's time is limited the large image/video is unavailable at the small device, or when the viewer must examine many images.

The Image Retargeting problem applies to video as well. Li et al. [12] and Wang et al. [21] both introduce a model to identify an ROI for each video frame, and use this to automatically crop the video. Our fisheye-view warping may apply to video retargeting.

A key step in automatic retargeting (and all of its variants described above) is to identify the important aspects of the image so that these can be emphasized. All methods (including the variant we introduce in Section 3.2) use some combination of high-level image understanding to identify objects that might be important and low-level visual salience to identify aspects of the image that the eye may be drawn to. The best known salience method is the framework of Itti et al. [7, 8] which has the advantage of being biologically based. Unfortunately, the approach is computationally demanding and difficult to tune, and therefore may be overkill for practical applications. Ma et al. [15] argue that heuristic methods that are computationally more efficient and simpler to implement are effective for retargeting problems despite not being a detailed model of the human visual system.

The use of generalized fisheye views to provide a central focus as well as context has a long history in the HCI and Information Visualization community dating back to 1982 by Spence and Apperley [18]. Since this pioneering work, researchers have proposed various applications, including calendar interfaces [2], menu interfaces [1], web image browsing [14], and map browsing [24].

Leung and Apperly [11] review “distortion oriented” display techniques, which refer to methods like ours that use non-linear transformations to create focus+context displays. Keahy [10, 9] provides a menagerie of transformations to achieve detail-in-context displays. Capendale and Montagnese [3] present another framework and survey how prior methods fit. Our work fits into these frameworks. The linear cartesian fisheye view technique of Section 4.3 is equivalent to the Bifocal Display [18]. The spline-based warps

we present are unlike any of the warps in the surveys in that they avoid distortions of a finite region and provide continuity and monotonicity guarantees. In applying these methods to the automatic retargeting problem we consider automatic control of the foveal region (ROI in our terminology).

Analysis of focus+context displays is helpful in understanding our methods. For example, Zanella et al. [25] explore the importance of making the distortions visible to the viewer, which our method does (distortion in photographs are usually obvious), while automatic cropping does not.

### 3 AUTOMATIC IMAGE RETARGETING

The image retargeting problem takes a *source* image and produces a *target* image that is better suited for the target display. For this paper, we consider retargeting problems where the source image is larger than the target. This common situation arises when the target is a portable device or a thumbnail. Often the change in size is also accompanied with a change in aspect ratio.

Because the target image is smaller than the source, some information is necessarily thrown away. The naïve approach to retargeting scales the source image to the target size with the appropriate downsampling. This uniformly throws away detail information across the image. The loss of detail can make important parts of the image difficult, or impossible to recognize. An aspect ratio change requires either non-uniform scaling (which “squishes” the image in distorting ways) or not filling the target image, which wastes the already small target image space.

The core problem with naïve retargeting is that by uniformly throwing away information, it ignores the fact that some parts of the image are more important than others. In a digital photograph, a face that covers a tenth of the image in each dimension is clear enough to be recognized: if this image is downsampled to the resolution of a cell phone, the size is so small that it is difficult to determine that there is a face, never mind identifying the person. Scaling treats the face as any other part of the image, disregarding the fact that it may be more important.

Intelligent retargeting methods attempt to make an informed choice as to what image information to throw away. If part of an image is determined to be important, that part should be kept (if possible) during retargeting, at the expense of less important parts of the image. An intelligent retargeting method would insure that the face in the example of the preceding paragraph remains large enough to be recognized by giving more space to the face and less space to other parts of the image. The most obvious way to do this is *cropping*, that is keeping the important piece and discarding the rest of the image.

Intelligent retargeting, therefore, consists of two parts: first, the important aspects of the image must be identified; second, a target image must be created that discards more information from the less important aspects of the image. An automated method for retargeting must provide automatic methods for both parts. Good methods for the second part of the process (producing the target image) are also valuable

in a semi-automatic setting where the important aspects of the source are manually annotated.

Truly understanding what is important in an image requires a thorough understanding of what the image contains and what the viewer needs. Fortunately, recent results [4, 15, 19] suggest that some heuristics work well (albeit imperfectly) on a broad class of images. The two heuristics that are used in most (if not all) image retargeting are: specifically recognizable objects (faces) are usually important; and regions of the image that are most likely to attract the low-level visual system are likely to be important. The remainder of this section describes how we have realized these heuristics and describe how we employ them to implement automatic cropping.

### 3.1 Importance Maps

Given no other information about the meaning of an image or the needs of the viewer, many have proposed visual saliency as an approximation of importance. The intuition for this heuristic is that the parts of the image that stand out or are likely to be noticed by the low-level human visual system are most likely to be important to the meaning of the overall image. A *saliency map* computes, for each pixel or image block, a value of how much it attracts the visual system.

The work of Itti and his colleagues [7, 8] provides a saliency metric that is based on the neuroscience of the human visual system. The method is computationally expensive and difficult to implement, although a publicly available implementation has made it a standard tool in automatic retargeting. Ma et al. [15] provide a simpler saliency metric based on heuristics observations of how people perceive images. Their experiments confirm its success at identifying the ROI for image retargeting and similar applications. Despite being efficient enough to implement on a cellular phone, their experiments suggest that their approach is at least as effective as Itti’s method [23]. Our experience is similar: our implementation of Ma’s contrast-based technique is more efficient and (by anecdotal evaluation) leads to better results for image retargeting than using an implementation of Itti’s method available from the author website<sup>1</sup>.

We use the following implementation of Ma’s contrast-based method [15] for determining the saliency map:

1. Transform images into a perceptually uniform color space (Lu\*v\*) [6].
2. Quantize colors uniformly into the range of [0 . . . 31].
3. Downsample images by 4 in each dimension.
4. Compute the saliency value  $S_{i,j}$  as the weighted sum of the contrast differences between the pixel  $(i, j)$  and each other pixel in its neighborhood. That is,

$$S_{i,j} = \sum_{q \in \Theta} w_{i,j} d(p_{i,j}, p_q) \quad (1)$$

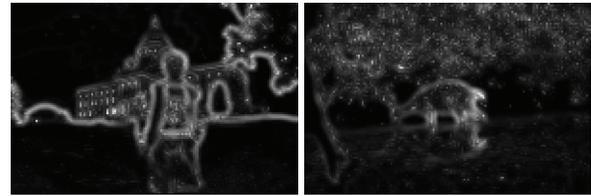
$$w_{i,j} = 1 - r_{i,j}/r_{max}$$

where  $\Theta$  is the neighborhood of  $(i, j)$  (a  $5 \times 5$  square in our implementation),  $p_{i,j}$  is the color of the pixel at  $i, j$  in

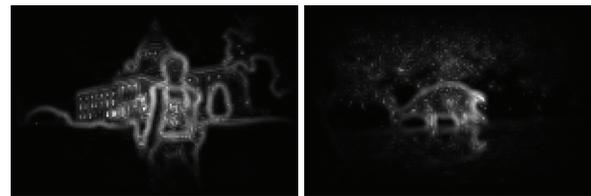
<sup>1</sup>Part of the difference may be due to the difficulties in tuning parameters in Itti’s method, particularly to deal with scale sensitivity.



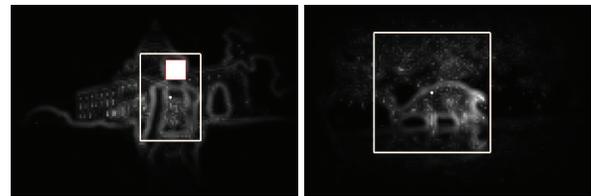
(a) Original image (w/ROI shown)



(b) Contrast map (non-center weighted)



(c) Saliency map (center weighted)



(d) Importance map (§3.2) w/ROI

Figure 1: ROI extraction. ROIs are indicated by thick rectangles, and faces are indicated by a solid rectangle. The values in each map are normalized to [0, 255] for the sake of demonstration.

the downsampled image, and  $d$  computes the magnitude of the difference between colors (we use the  $L^2$  norm). The weighting factor  $w_{i,j}$  is used to account for the heuristic that the center of an image is usually more visually salient.  $r_{i,j}$  is the distance from  $(i, j)$  to the image center and  $r_{max}$  is the biggest distance to the image center.

Examples of saliency maps are shown in Figure 1(b) and (c).

Another heuristic for visual importance in images is that certain identifiable objects are often important. Most automatic image retargeting methods use face detection because faces are almost always important and are a common feature in images. Other detectors, such as the text detector used in [4], could be added.

In practice, automatic retargeting needs to combine low level visual saliency, which may miss important objects, with specific object detectors, which may not apply to all images. Our implementation combines a face detector using the Adaboost

method [20] with the implementation of the contrast-based saliency metric described above.

### 3.2 Region of Interest Finding

Many automatic retargeting methods require the importance information to be collected as a discrete region of interest (ROI). The method we used, inspired by that of Suh et al. [19], uses an *importance map* that combines saliency and object information to create a single, rectangular ROI.

The ROI is defined as a rectangle with minimal area that contains both “enough importance” and the identified important objects. Enough saliency is defined as when the sum of the importance for all pixels in the rectangle is a large enough proportion (a parameter  $\lambda$ , set to 0.7 in our experiments<sup>2</sup>, except where noted) of the total importance across the entire image.

Rather than using an exhaustive search to find the best possible ROI, we use a greedy algorithm. Given the saliency map and a list of rectangles containing faces, our method finds an initial candidate ROI and grows it until it contains sufficient importance. It works as follows:

- Step 1: Importance map initialization: pixels in the importance map that are part of faces are set to have the maximum value appearing in the saliency map, and the others are set to the corresponding saliency value. The total saliency of the image is also computed.
- Step 2: ROI initialization. If only one face is detected, we fix the center of the face area, magnify it 1.3 times both in width and height, and take this as the initial ROI. In the case of multiple faces, we select the dominant one. We define the dominant face as the one with the largest value of the product of the face’s area, its centrality (using the weight  $w_{ij}$  from above), and the pose weight of the face (giving higher weight to frontal views than profile views). If no face is detected, we exhaustively search for a small window of size  $20 \times 20$  that contains most saliency.
- Step 3: ROI growing. If the ROI contains less than  $\lambda$  percent of the total importance value, we grow the ROI in one of the four directions, left, right, top, and bottom. We select the direction where growing can increase the mean importance value in the ROI most significantly. This step is repeated until the ROI contains sufficient saliency.

Some examples of ROI computation are shown in Figure 1.

### 3.3 Automatic Cropping

Once a rectangular ROI is determined, it is easily extended to an automatic cropping method by first growing the ROI in one direction so that it has the same aspect ratio as the target image size. This updated ROI is then cropped from the source image and scaled to the target image size. We use this method to provide the automatic cropping technique used in the evaluations of Section 5.

The automatic cropping method we provide is similar to the successful methods of previous papers. Our method differs

<sup>2</sup>We have determined  $\lambda$  empirically. The automatic approach of [19] does not apply to center-weighted importance maps, and our experiments with automatic parameter finding have not produced results that we feel are superior to the fixed, well-chosen value.

in some of the details of how the ROIs are computed, and that we use a fixed value for  $\lambda$ . While we do not believe these differences to be significant, other ROI computation techniques could be used with the warp-based retargeting methods of the next section.

All automatic cropping methods, regardless of how they compute the ROI, share a significant drawback: they completely discard the parts of the image that are not in the ROI. This is a significant drawback because the less important aspects of the image may still be important, if only to provide context for the more important parts. Also, in the event that ROI finding fails to correctly identify an important object (which is inevitable since it is a heuristic process), the failure is catastrophic since the important aspects of the image are discarded. These issues motivated us to create a new image retargeting method that employs ROI finding but does not completely discard the remainder of the image.

## 4 FISHEYE-VIEW WARPING

Given the ROI for an image, for example obtained by the method of the previous section or by manual annotation, our goal is to define a method for obtaining a target image that provides emphasis to the ROI without completely discarding other parts of the image. We view this transformation as an image warp. Image warps are conveniently described by a warping function that maps positions in the source image to positions in the target image:

$$x', y' = \mathbf{f}(x, y).$$

Once the warping function is defined, warping can be performed by image resampling. While this step must be performed correctly, resampling is a well-developed science, see Wolberg’s book [22] for a tutorial. The resampling methods we use are detailed in Section 4.4.

Scaling is a trivial case of warping where

$$\mathbf{f}(x, y) = \left( \frac{dst_x}{src_x} x, \frac{dst_y}{src_y} y \right),$$

where  $src_x, src_y$  is the size of the source image and  $dst_x, dst_y$  is the size of the destination image.

Note that the derivative of the warping function gives the magnification of the target at a particular point. For scaling, this magnification is constant - it is a *linear* warp. To create an image retargeting warp, we need to create warping functions that can provide different magnifications in different parts of the image, and are therefore, non-linear.

To avoid obscuring a portion of the image, the warping function must not “fold over” itself. This means that the warping function must be monotonic.

Some warping functions can be written as independent functions of one variable:

$$x' = f_1(x), y' = f_2(y).$$

We call these one dimensional functions *scaling functions* and prefer them because they facilitate reasoning about the derivatives (for example to insure monotonicity). All of the warps discussed in this paper can be expressed this way.

#### 4.1 ROI Position and Size

Our first step in deciding on how to do a warp-based image retargeting is to determine the position and size of the ROI in the target image. To insure the clarity of the ROI, we choose to preserve its aspect ratio, so we must choose a single magnification factor,  $\kappa$ .

We define  $\kappa_{max}$  to be the maximum possible magnification of the ROI, that is, the magnification that causes the ROI to completely fill the target image. It is computed as

$$\kappa_{max} = \min\left(\frac{dst_x}{roi_x}, \frac{dst_y}{roi_y}\right).$$

We choose  $\kappa$  to fill a specified percentage of the target with the ROI. Our rationale is that the ROI is defined to contain at least a specified percentage ( $\lambda$  in Section 3) of image importance. For all of the experiments in this paper,  $\kappa = .70\kappa_{max}$ . We bound the value of  $\kappa$  to be at most 1, since we do not want to enlarge the ROI beyond its original size.

Given the new ROI size, a new position is determined so that the proportions of the border remain constant. For example, the ratio of the amount of the image to the left of the ROI and the right of the ROI are the same in both the source and target image.

#### 4.2 Radial Warping

A photographic tool for emphasizing a portion of the image is the fisheye lens. Distortion-free extremely wide-angle optics are difficult to produce, and impossible to produce for fields of view greater than  $180^\circ$ . Therefore, photographers use lenses that project onto curved (usually spherical) surfaces since they provide well-characterized distortions. These devices, known as fisheye lenses, produce distinctive images where most straight lines appear curved. Another characteristic of fisheye imagery is that objects closer to the center<sup>3</sup> of the image appear larger than those further away from the center point.

The radial distortion of a fisheye lens makes the size of an object depend on its distance from the ‘‘center’’ or focus point. This is easily described by using polar coordinates with the origin at the center and a scaling function for the radius:

$$\theta' = \theta, \quad r' = f(r).$$

If  $f$  is linear, there is no distortion of the image (the warp is equivalent to uniform scaling). To achieve the fisheye effect,  $f(r)$  must be monotonically increasing, but its derivative must be non-increasing as  $r$  increases.

To use the radial fisheye for image warping, we define the center to be the center of the ROI, and create an appropriate scaling function. For simplicity in design and description, we define the scaling function as a curve in two dimensions. The curve must interpolate the point  $(0, 0)$  so the source center maps to the target center,  $(r_{roi}, r'_{roi})$  so that the boundary of the ROI maps to the correct location, and  $r_{max}, r'_{max}$  so the boundary of the images maps correctly. If the ROI and

<sup>3</sup>Where the optical center of the lens projects onto the image, which may not necessarily be the center of the image.

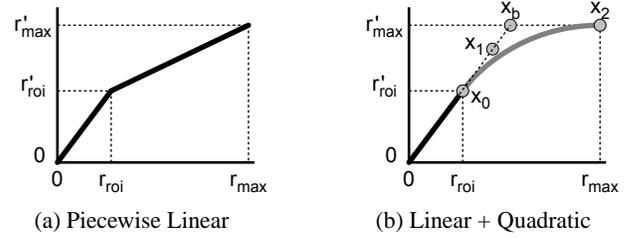


Figure 2: Scaling functions for radial fisheye warps.

images are not circles, these values will be different for varying values of  $\theta$ .

The scaling function interpolates the three points in a piecewise fashion. Within the ROI, it linearly interpolates between  $(0, 0)$  and  $(r_{roi}, r'_{roi})$  to provide uniform scaling. A simple scaling function also uses linear interpolation between  $(r_{roi}, r'_{roi})$  and  $(r_{max}, r'_{max})$ . This does not provide the effect of objects further from the center (or ROI) being smaller, and has a discontinuity in the magnification. To address these issues, we use a quadratic function, specified in Bézier form. These functions are shown in Figure 2.

To use a quadratic Bézier curve for the second piece of the scaling function, its endpoints are given as  $\mathbf{x}_0 = (r_{roi}, r'_{roi})$  and  $\mathbf{x}_2 = (r_{max}, r'_{max})$ . We must determine the position of its middle control point,  $\mathbf{x}_1$ . Our desire to have the initial slope of the curve match the slope of the line segment dictates a line that the point must lie on. Properties of Bézier curves [17] allow our monotonicity requirement to be implemented by a constraint that the point be to the left of and below its endpoint  $\mathbf{x}_2$ . These limit the placement of the middle control point to lie along a line segment, introducing a new parameter  $\alpha$  to determine where exactly the point lies. The geometry of this is shown in Figure 2(b). We define the point  $\mathbf{x}_b$  as the maximal possible value for the points position, that is where the line with slope  $\kappa$  intersects the line where  $r' = r'_{max}$ . The middle control point is then

$$\mathbf{x}_1 = (1 - \alpha)\mathbf{x}_0 + \alpha\mathbf{x}_b.$$

The parameter  $\alpha$  dictates how much emphasis is placed on areas near the ROI. When  $\alpha = 1$ , pixels at the source image edge have zero magnification, so pixels near the edge have very small magnification, leaving more space for pixels near the ROI. Smaller values of  $\alpha$  provide a more uniform distribution of the space, giving more space to the pixels near the edge. At the extreme,  $\alpha = 0$ , the curve becomes a line segment so all pixels receive the same magnification.

Although the exact choice of  $\alpha$  is a minor detail, we have chosen to tune it based on image content. The intuition for our heuristic is that the more information that exists at the periphery of the image, the smaller  $\alpha$  should be. We use the following equation to compute  $\alpha$  based on the distribution of importance map values:

$$\alpha = 1 - \frac{\sum_{(r,\theta) \in Context} r A_{r,\theta}}{r_{max} \sum_{(r,\theta) \in Context} A_{r,\theta}},$$

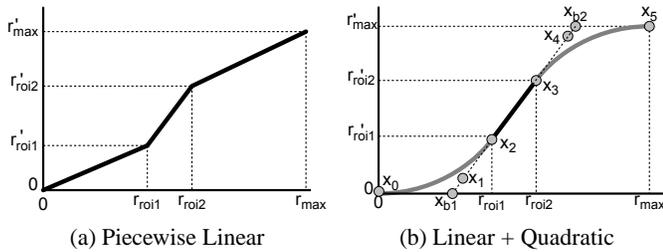


Figure 3: Scaling functions for cartesian fisheye warps.

where *Context* is the set of pixels not in the ROI, and  $A$  is the importance map as in Section 3.

Note that we have defined the scaling function as a spline in two dimensions. This was required to give us sufficient flexibility in the control of the derivatives despite using a low-order polynomial. To use the curve to evaluate  $r' = f(r)$  we must solve the equation numerically. Because the curve is monotonic and smooth, this is very efficient.

### 4.3 Cartesian Fisheye Warping

In practice, we find that radial warps create images that are unpleasantly distorted as shown in Figure 4. The radial nature of the warping function causes bending that is very noticeable, especially in scenes with straight lines. We have therefore used a variant of fisheye view warping that creates less bending. We call the method *Cartesian Fisheye Warping* because it applies the fisheye view warp scaling function independently in each cartesian dimension.

To understand the method, consider that identifying a rectangular ROI effectively divides the image into nine pieces (a  $3 \times 3$  grid). Given that the transformations of these nine pieces are known (since the methods of Section 4.1 provide the size and position of the central ROI piece), we can scale each piece independently so that it fits in its target location. We call the method that scales each piece linearly a *Linear Cartesian Fisheye warp*.

The Linear Cartesian Fisheye warp can be viewed as the application of a 3 piece, piecewise linear scaling function in each dimension, as shown in Figure 3(a). Just as in the radial case, we can replace the scaling functions outside of the ROI with quadratic splines to provide continuity in the magnification and so that objects farther from the ROI are smaller in the final image. All of the spline design from the previous section is applied, except that it is applied symmetrically to the beginning and end of the curve, as shown in Figure 3(b). Figure 4 shows examples of cartesian fisheye warps for both linear and quadratic scaling functions.

### 4.4 Implementation Details

The warping function provides a continuous mapping from positions in the source image to the target image. Resampling is implemented using the texture mapping operations provided by graphics hardware. Specifically, we map a source grid, with each element  $10 \times 10$  pixels, to their target positions. This forward mapping warp method approximates the non-linear deformations with a piecewise bilinear function.

Texture mapping provides an easy way to implement an efficient resampling method. However, most texture mapping provides only isotropic filtering, which may cause unnecessary blurring in areas where pixels are stretched non-uniformly. In practice, this does not seem to have a major impact on the visual quality of our results: the portions of the images that receive anisotropic transformations are the periphery which is already quite distorted by the transformation itself. To provide better performance on the anisotropic filtering, we use texture mapping to produce a target image twice the size of the final target, and use high quality image scaling (bicubic filtering in Photoshop) to produce the final image.

Source images larger than 1000 pixels on a side are down-sampled (typically to have the longest dimension be 800 pixels) before retargeting. This is done to speed ROI finding.

The time needed for retargeting each image contains two parts, time for ROI finding and time for warping. Warping using graphics hardware is fast: our method supports interactively adapting a  $1600 \times 1200$  image with real time updates while warp parameters are changed with a slider. However, automatic ROI finding is more time consuming. Timing depends on image size and content. For  $800 \times 600$  images, some may take as long as 20 seconds, although most are faster. Timings were made on a desktop PC with 1GB memory and a 2GHz Intel Pentium 4 CPU. We view image retargeting as a task better suited to servers than to the mobile devices themselves.

## 5 EVALUATION

In principle, the fisheye view warping methods of the last section meet our goals for automatic image retargeting by providing a method that emphasizes the important aspects of the image without discarding other parts. Our initial experience with the prototype implementation lead us to believe that the approach is successful. Deciding amongst the warp variants of the previous section is subjective (examples are shown in Figure 4).

We have conducted two empirical studies of our methods to gain a better assessment of it. First, a user study explored the subjective aspect of the work: which of the particular warps people prefer, and how our method compares to previous ones. A second study aimed to assess the methods robustness both absolutely, and relative to automatic cropping.

Our preliminary studies are subjective. Part of this is that our initial concern were the aesthetic issues in creating distorted images. Also, objectively measurable tasks are not clearly defined for small device image display (as they are for other applications like image database browsing [19]).

### 5.1 User Study

We conducted a preliminary user study to assess viewer reaction to the warped images. Our goal was to determine which of the warping variants is preferred, and to determine if users are willing to accept the distortions.

We selected 35 images from images taken by members of our research group. All source images came from multi-megapixel digital cameras and had either a  $4 \times 3$  or  $3 \times 4$  aspect

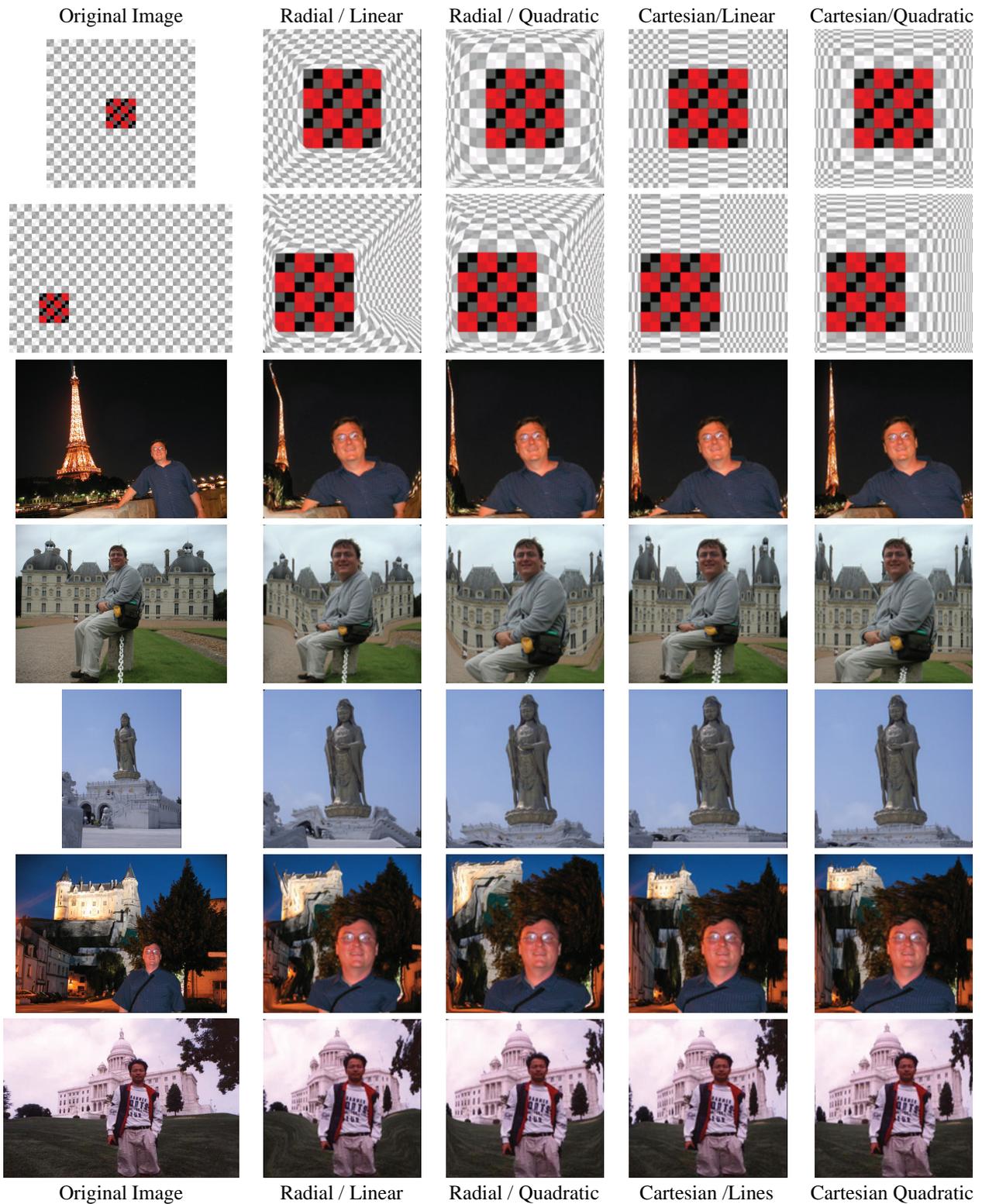


Figure 4: Examples of the variants of fisheye warps for image retargeting. Original images are uniformly scaled to fit. ROIs for the checkerboard images were specified manually. Target size is  $160 \times 160$ . The top two images were of size  $400 \times 400$  and  $600 \times 400$ , and had their ROIs and parameter values specified manually.



Figure 5: Example of web page display for user study. Subjects were shown two versions of each image on simulated cell phones.

ratio. We selected images for which the ROI finding process was successful, and made slight, manual adjustments to the ROI of 3 images to prevent failures from automatic cropping.

For each trial, a source image was retargeted using two different methods to a common size for a cell phone display (some to  $160 \times 160$ , others to  $128 \times 192$ ). The same ROI was used for both methods. Retargeted images were prepared in advance, and the same ones were used for all subjects. The order of presentation was randomized independently for each subject, as was which of the two methods appeared on the left or right in each trial. Subjects were shown the two different retargeted images side by side as simulated cell phone displays on web pages, as in Figure 5. For the first 28 trials subjects were asked to select the “image they would prefer to receive.” Subjects were forced to choose between the two target images. A prior pilot study on a dozen people showed similar results to the final study.

The user study was conducted via the web. Subjects voluntarily responded to advertisements posted to mailing lists and were not compensated for their time. To our knowledge, the subjects did not know about our work. We obtained 45 responses from a pool of computer science graduate students in our department (excluding members of our research group), and 14 responses from a pool of graduate students of varying disciplines. The result from each subject pool was similar, so we combine them here. Our human subjects approval precluded us asking any identifying or demographic information of the subjects.

Trials compared spline-based cartesian fisheye (SCF) warping to naïve scaling, automatic cropping, spline-based radial warping, and a “bilinear” warping<sup>4</sup>. For each trial type, there

<sup>4</sup>The bilinear warping uses a constant scale for each of the 9 cartesian regions. Because of its simplicity and inferior performance we have not dis-

	images	mean	std. dev.	p
Phase 1				
scale	7	6.63 (95%)	0.76	0.00002
crop	10	6.73 (67%)	2.12	0.21
radial warp	5	4.14 (83%)	1.04	0.058
bilinear	6	5.54 (92%)	.8371	.001
Phase 2				
inform	7	5.73 (82%)	1.45	.062

Table 1: Summary of user tests explained in text.

were a different number of trials (7, 10, 5, and 6 respectively). To assess each type, we counted the number of trials where the SCF image was selected. On average, subjects chose an SCF image over a scaled image 6.63 out of 7 times (95%), an SCF image over an auto-cropped image 6.73 out of 10 times (67%), and an SCF image over a radial warped image 4.14 out of 5 times (83%). Each of these results suggest that SCF is preferred to the other methods, however the variance for the latter two is high. We computed significance using a *t* test ( $n = 59$ ) to provide the probability of the population mean showing no preference for SCF despite the sample mean showing it. These results are summarized in Table 1.

An interesting result of this study is that our subjects did not prefer undistorted automatic cropping<sup>5</sup> to cartesian fish-eye warps. In fact, our sample shows a preference for the warped images, although it is statistically insignificant. This result is surprising because we did not explain the benefits of the warped images to the subjects, they simply were shown distorted images. Several of the image pairs shown to viewers are shown in Figure 6.

A second phase of the user study asked viewers which image version they found “more informative.” Subjects selected the fisheye view an average of 5.73 times from 7 examples (82%), with a standard deviation of 1.45. Confidence that the population prefers the fisheye version is .062 (t-Test,  $n = 59$ ). Some of this variance can be attributed to differing interpretations of the wording “more informative.”

While the results of our preliminary user study are inconclusive (except to re-confirm the result of [19] and [4] that scaling is a poor choice for image retargeting), we find them encouraging. Although subjects do not necessarily prefer distorted images, they do not seem to reject them either. This makes it plausible that viewers will be willing to accept distorted images in order to receive their benefits.

## 5.2 Evaluation of Robustness

Given the heuristic nature of ROI finding, we were concerned about the robustness of automatic image retargeting. That is, given a new image, how confident can we be that an automatic retargeting method gives an acceptable result. Our goal for this evaluation was both to get some sense of how our fisheye view warping method fares in an absolute sense and also to empirically support our claim that fisheye warp-

—  
cussed its implementation.

<sup>5</sup>We emphasize that this is not a test of any specific previous automatic cropping algorithm as our implementation §3.3 combines elements of what we consider to be the two state of the art approaches.



Figure 6: Comparisons of automatic cropping (left of each pair) and linear cartesian warping (right of each pair).

ing has better robustness properties than automatic cropping. Unfortunately, since “success” in retargeting is subjective, and the realm of images is virtually infinite and diverse, a true robustness test is difficult. The following experiment tries provide some answers to our robustness questions.

A fair test requires a large set of images that were obtained in an unbiased manner (e.g. not taken by, or hand selected by us). Therefore, we obtained a large commercial image library of 55720 images. This library is the same “Corbis” library used by Suh et al. [19] in their study<sup>6</sup>. This library contained a wide variety of subject matter, ranging from texture tiles to natural scenes to studio photographs of individual objects. These images were of varying size, ranging from 100-256 pixels on a side, with a variety of aspect ratios ranging from 0.39-2.56.

From the library of 55720 images we created a subset of 261 images by random sampling without replacement. 20 images were discarded as inappropriate, for example if the image was an abstract pattern. The 241 images were run through our automatic retargeting system to produce target images of a size of 100×100. All images were processed using both automatic cropping algorithm (§3.3) and Cartesian Spline Fish-eye (§4.3). The same ROI computation was used for both methods, and we tested varying values of  $\lambda$ . Results were manually categorized as follows:

**Failure:** We consider a retargeting a failure if the result was clearly missing an important aspect of the source image. For example, if a picture of two people only shows one.

**Success:** We consider a retargeting a success if the result conveys the same message as the original image.

This categorization is subjective and was done by the authors.

The results of the tests are summarized in Table 2. Fish-eye view warping rarely fails, even at small values of  $\lambda$ . Automatic cropping fails more often, even at higher  $\lambda$  values.

<sup>6</sup>Licensing terms of the image set preclude showing them in this paper.

$\lambda$	automatic cropping			spline cart. fisheye		
	.6	.7	.8	.6	.7	.8
Fail	91	86	39	10	10	5
Succeed	150	155	202	231	231	236

Table 2: Summary of robustness test explained in text.

In practice, smaller  $\lambda$  values are necessary for warping to achieve equivalent magnification of the ROI. Even adjusting for this difference, warping provides better robustness.

For automatic cropping, the high failure rate may be partially attributed to the small sizes of the source images. Face detection and saliency map generation work better when given more data. Fisheye retargeting provides better robustness given the poor ROI finding.

Even allowing for experimenter bias, we feel these results suggest that fisheye warping is more robust than automatic cropping given the same ROI computation. Some specific examples of where fisheye warping succeeded while automatic cropping failed include:<sup>6</sup>

- The images “doctor” and “photographer” were cropped tightly around a person’s face. The warped image show the stethoscope and camera (albeit distorted).
- The images “black and white” and “mother and child” were cropped to show a single person. In the warped image, The second face was distorted, but still present.
- Images of “surfer” and “tennis player” cropped the images to remove the person’s head.
- “Child blowing out candles” was cropped to make us wonder what he was doing. The warped image included a birthday cake, albeit squished.

## 6 DISCUSSION AND CONCLUSION

This paper introduces a new method for automatic image retargeting that uses fisheye view warps. The method emphasizes important aspects of images without completely discarding the remaining parts. Therefore, it can provide more informative images than prior methods as well as better robustness.

Many aspects of our methods can be improved. In particular, better methods for determining the ROI, and for identifying the important aspects of the context will be useful. More content- and context-sensitive methods will allow better determination of what can and cannot be distorted. Understanding the minimal perceptible size of identified features can help prevent over-emphasizing the ROI or under-representing other image aspects.

More thorough evaluation of the technique is important. A key goal is to assess the suitability and effectiveness of the methods on real tasks involving either real mobile devices or database browsing problems. Our preliminary results indicate users like the additional information in the retargeted results, however we have no proof that they can make use of the information that is available in distorted form at the periphery of the warped results. It is our belief that even if this peripheral information is too distorted to be recognized, it still provides indication that *something* is in the original image. User studies will also be important to help us tune the subjective aspects of our methods.

The fundamental limitation of our method comes from the assumption that the image has a single ROI, and that conveying this ROI clearly is the most important task of the target image. While our method does not fail as catastrophically as automatic cropping when these assumptions are not met, alternative methods that address specifically multi-ROI retargeting may provide better results in these cases.

## 7 ACKNOWLEDGMENTS

This work was supported by NSF grants IIS-0097456 and IIS-0416284. We thank Yu-Chi Lai, Tom Brunet, Ben Bederson, and Bruce Gooch for their help with this project.

## References

- [1] Benjamin Bederson. Fisheye menus. In *Proceedings UIST '00*, pages 217–225, 2000.
- [2] Benjamin Bederson, Aaron Clamage, Mary Czerwinski, and George Robertson. Datelens: A fisheye calendar interface for PDAs. *ACM Trans. Comput.-Hum. Interact.*, 11(1):90–119, 2004.
- [3] M. S. T. Carpendale and Catherine Montagnese. A framework for unifying presentation space. In *Proceedings of UIST '01*, pages 82–92, 2001.
- [4] Li-Qun Chen, Xing Xie, Xin Fan, Wei-Ying Ma, Hong-Jiang Zhang, and He-Qin Zhou. A visual attention model for adapting images on small displays. *ACM Multimedia Systems Journal*, pages 353–364, 2003.
- [5] Xin Fan, Xing Xie, He-Qin Zhou, and Wei-Ying Ma. Looking into video frames on small displays. In *Proceedings ACM Multimedia 2003*, 2003. Short Paper.
- [6] J. Foley, Van Dam A., S. Feiner, and J. Hughes. *Computer Graphics: Principles and Practice, 2nd edition*. Addison Wesley, 1997.
- [7] L. Itti and C. Koch. Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2(3):194–203, Mar 2001.
- [8] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(11):1254–1259, 1998.
- [9] T. Alan Keahey. The generalized detail-in-context problem. In *Proceedings IEEE Symposium on Information Visualization*, 1998.
- [10] T. Alan Keahey and Edward Robertson. Techniques for non-linear magnification transformations. In *Proceedings IEEE Symposium on Information Visualization*, 1996.
- [11] Y. K. Leung and M. D. Apperley. A review and taxonomy of distortion-oriented presentation techniques. *ACM Trans. Comput.-Hum. Interact.*, 1(2):126–160, 1994.
- [12] Ying Li, Yu-Fei Ma, and Hong-Jiang Zhang. Salient region detection and tracking in video. In *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)*, 2003.
- [13] Hao Liu, Xing Xie, Wei-Ying Ma, and Hong-Jiang Zhang. Automatic browsing of large pictures on mobile devices. In *11th ACM International Conference on Multimedia, Berkeley*, 2003.
- [14] Hao Liu, Xing Xie, Xiaou Tang, Zhi-Wei Li, and Wei-Ying Ma. Effective browsing of web image search results. In *MIR '04: Proceedings of the 6th ACM SIGMM workshop on Multimedia information retrieval*, pages 84–90, 2004.
- [15] Yu-Fei Ma and Hong-Jiang Zhang. Contrast-based image attention analysis by using fuzzy growing. In *Proceedings ACM Multimedia 2003*, pages 374–381, 2003.
- [16] Vidya Setlur, Saeko Takagi, Ramesh Raskar, Michael Gleicher, and Bruce Gooch. Automatic image retargeting. In *Technical Sketch, Sigraph 2004*, 2004.
- [17] Peter Shirley. *Fundamentals of Computer Graphics*. AK Peters, 2002.
- [18] R. Spence and M. Apperley. Data base navigation: an office environment for the professional. *Behaviour & Information Technology*, 1(1):43–54, 1982.
- [19] Bongwon Suh, Haibin Ling, Benjamin B. Bederson, and David W. Jacobs. Automatic thumbnail cropping and its effectiveness. In *Proceedings UIST '03*, pages 95–104, 2003.
- [20] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proc. Conf. on Computer Vision and Pattern Recognition*, pages 511–518, 2001.
- [21] Jun Wang, Marcel Reinders, Reginald Lagendijk, Jasper Lindenbergh, and Mohan Kankanhalli. Video content presentation on tiny devices. In *IEEE International Conference on Multimedia and Expo (ICME 2004)*, 2004. Short Paper.
- [22] George Wolberg. *Digital Image Warping*. IEEE Press, 1990.
- [23] Xing Xie and Yu-Fei Ma. Personal communication, November 2004.
- [24] Christopher C. Yang, Hsinchun Chen, and K. K. Hong. Visualization tools for self-organizing maps. In *DL '99: Proceedings of the fourth ACM conference on Digital libraries*, pages 258–259, 1999.
- [25] A. Zanella, M. S. T. Carpendale, and M. Rounding. On the effects of viewing cues in comprehending distortions. In *Proceedings of ACM Nordi-CHI '02*, pages 119–128, 2002.