Rochester Institute of Technology

RIT Scholar Works

Theses

9-27-2006

Zoom techniques for achieving scale invariant object tracking in real-time active vision systems

Eric Nelson

Follow this and additional works at: <https://scholarworks.rit.edu/theses>

Recommended Citation

Nelson, Eric, "Zoom techniques for achieving scale invariant object tracking in real-time active vision systems" (2006). Thesis. Rochester Institute of Technology. Accessed from

This Thesis is brought to you for free and open access by RIT Scholar Works. It has been accepted for inclusion in Theses by an authorized administrator of RIT Scholar Works. For more information, please contact ritscholarworks@rit.edu.

Zoom Techniques for Achieving Scale Invariant Object Tracking in Real-time Active Vision Systems

by

Eric D. Nelson

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Computer Engineering

Supervised by

Associate Professor Dr. Juan C. Cockburn
Department of Computer Engineering
Kate Gleason College of Engineering
Rochester Institute of Technology
Rochester, New York
July 2006

Approved By:

Dr. Juan C. Cockburn
Associate Professor
Primary Adviser

Dr. Andreas Savakis
Professor and Department Head, Department of Computer Engineering

Dr. Roy Czernikowski
Professor, Department of Computer Engineering

Thesis Release Permission Form

Rochester Institute of Technology
Kate Gleason College of Engineering

Title: Zoom Techniques for Achieving Scale Invariant Object Tracking
in Real-time Active Vision Systems

I, Eric D. Nelson, hereby grant permission to the Wallace Memorial Library to reproduce my thesis in whole or part.

Eric D. Nelson

Date

Dedication

I would like to dedicate this work to my Dad, for all his encouragement, support, and wisdom. And also to the memories of my Grandma Kozlow and Uncle Bob.

Acknowledgments

For and foremost, I would like to thank my advisor Dr. Juan C. Cockburn serving as a patient mentor throughout this thesis. Also I would like to thank my committee, Dr. Roy S. Czernikowski and Dr. Andreas Savakis. Many thanks to Jared Holsopple, John Ruppert, Justin Hnatow, Yuriy Luzanov, Eugene Pokhis, and Amish Rughoonundon for developing a code base from which to start my work. Thanks to Thomas Cockburn for lending me his R/C car. Much gratitude for the always awesome Laura Lamp who demanded upon the improvement of my writing style by sacrificing many of her weekends to meticulously edit the paper. Various thanks to: Rick Tolleson, Sergey Katsev, Evan Clark, Jared Holsopple, Jenelle Penders, and Mark Seidman.

Abstract

In a surveillance system, a camera operator follows an object of interest by moving the camera, then gains additional information about the object by zooming. As the active vision field advances, the ability to automate such a system is nearing fruition. One hurdle limiting the use of object recognition algorithms in real-time systems is the quality of captured imagery; recognition algorithms often have strict scale and position requirements where if those parameters are not met, the performance rapidly degrades to failure. The ability of an automatic fixation system to capture quality video of an accelerating target is directly related to the response time of the mechanical pan, tilt, and zoom platform—however the price of such a platform rises with its performance. The goal of this work is to create a system that provides scale-invariant tracking using inexpensive off-the-shelf components.

Since optical zoom acts as a measurement gain, amplifying both resolution and tracking error, a new second camera with fixed focal length assists the zooming camera if it loses fixation—effectively clipping error. Furthermore, digital zoom adjusts the captured image to ensure position and scale invariance for the higher-level application. The implemented system uses two Sony EVI-D100 cameras on a 2.8GHz Dual Pentium Xeon PC. This work presents experiments to exhibit the effectiveness of the system.

Contents

Dedication	iii
Acknowledgments	iv
Abstract	v
Glossary	xi
1 Introduction	1
1.1 Active vision	1
1.2 Thesis outline	2
2 Background Theory and Related Work	4
2.1 An introduction to cameras	4
2.1.1 Perspective projection	6
2.1.2 Weak perspective projection	7
2.1.3 Thin lens model	9
2.1.4 Calibration	10
2.2 Tracking	11
2.2.1 Classification and clustering via color segmentation	12
2.3 Focal length selection for optical zoom	13
2.3.1 A steady-state Kalman filter	13
2.3.2 Zoom invariant Kalman filter	15
2.3.3 Controlling focal length	23
2.4 Center of expansion	26
2.5 Summary	28
3 Dual Camera Zoom	30
3.1 Introduction	30
3.2 Recovering fixation	31

3.2.1	Camera view correspondence	31
3.2.2	Assisted control	33
3.2.3	Control arbitration	35
3.3	The experimental system	36
3.3.1	Pan and tilt	37
3.3.2	Zoom	38
3.3.3	System organization	43
3.4	Experiments on hardware	43
3.4.1	Pendulum experiment	43
3.4.2	Car experiment	47
3.5	Summary	50
4	Digital Scale Invariance	51
4.1	Choosing magnification for digital zoom	51
4.2	Fixating with digital zoom	53
4.3	Noise analysis	53
4.3.1	Impact of image noise	53
4.3.2	Simulations	54
4.4	Display arbitration	60
4.5	System Organization	62
4.6	Experiments on hardware	62
4.6.1	Pendulum experiment	62
4.6.2	Car experiment	67
4.7	Summary	71
5	Conclusion	72
5.1	Summary of this work	72
5.2	Future work	72
5.2.1	Advanced tracking	73
5.2.2	More cameras	73
5.2.3	Advanced zooming camera	73
A	Miscellaneous	74
A.1	Derivation of constant velocity Kalman Filter	74
	Bibliography	76

List of Figures

2.1	Pinhole camera [12]	4
2.2	Object size and object distance affect image size [12]	5
2.3	Perspective Projection [12]	6
2.4	Weak perspective projection [12]	7
2.5	Lenses allow more light to enter the pinhole [12]	8
2.6	Thin lens model	9
2.7	Rotational calibration	11
2.8	Image used to train color segmentation tracker	12
2.9	Object motion used in figures 2.10-2.14 [30]	17
2.10	Tracking with fixed focal length [30]	18
2.11	Tracking with variable focal length [30]	19
2.12	Tracking with variable focal length, with $R \propto f^2$ [30]	20
2.13	Tracking with variable focal length, with $Q \propto \frac{1}{f^2}$ [30]	21
2.14	Tracking with variable focal length, with $R \propto f^2$ and $Q \propto \frac{1}{f^2}$ [30]	22
2.15	The effect of varying the “forgetting factor” γ [30]	25
2.16	Example of center of expansion	27
3.1	Camera correspondence	32
3.2	Camera correspondence overhead view	33
3.3	Sony EVI-D100	36
3.4	Step response for pan and tilt	37
3.5	Block diagram for pan and tilt	38
3.6	Zoom reading step response, lookup table, and magnification step response	39
3.7	Pseudocode to create magnification lookup table	40
3.8	Zoom reading calibration	41
3.9	Control arbitration block diagram	42
3.10	Snapshots of pendulum experiment.	45
3.11	Pendulum results	46
3.12	Snapshots of car experiment.	48

3.13	Car results	49
4.1	Spinning coin problem	52
4.2	Unfiltered digital zoom	56
4.3	Kalman filter of digital zoom with $R_k = 10^{-4}$	57
4.4	Kalman filter of digital zoom with $R_k = 10^{-8}$	58
4.5	Kalman filter of digital zoom with $R_k = \frac{R}{h'_m{}^2}$	59
4.6	Digital zoom block diagram	61
4.7	Scale invariance for panoramic camera	63
4.8	Digital fixation for panoramic camera	63
4.9	Scale invariance for zooming camera	64
4.10	Digital fixation for zooming camera	65
4.11	Display arbitration pendulum experiment	66
4.12	Scale invariance for panoramic camera	67
4.13	Digital fixation of panoramic camera	68
4.14	Scale invariance of zooming camera	68
4.15	Digital fixation of zooming camera	69
4.16	Display arbitration car experiment	70

List of Tables

4.1	Filter comparison in presence of image noise	55
-----	--	----

Glossary

A

assisted control When a camera can only reliably fixate upon an object with both information acquired from its own sensors, along with information acquired from other cameras' sensors. Needed when the target object falls out of the camera's view. See autonomous control.

autonomous control When a camera can reliably fixate upon an object with information acquired only from its own sensors. Requires that the target object be in view of the camera. See assisted control.

C

calibration When various intrinsic and extrinsic properties of a lens are known, then the lens is considered calibrated. Angles and distances can be calculated with high accuracy in images acquired from a calibrated lens. In this work, calibration is limited to only knowing how many pixels in a captured image correspond to a degree on the camera's pan or tilt axis. This partial calibration is low maintenance, but is obviously not as useful as full calibration where every parameter is known.

Cartesian Theater A metaphysical theater in the brain where reason and the senses, including vision, come together.

center of expansion The point in the image around which zoom happens; all images appear to expand around this point when increasing magnification. The point

is generally fixed for optical zoom while it is flexible for digital zoom. See section 2.4.

control arbitration The process that decides whether to allow autonomous or assisted control for the zooming camera.

D

digital zoom When the size of an image is increased using upconversion and/or down-conversion. The target object's detail does not increase. For large magnifications, the resulting image can appear very pixelated. Compare to optical zoom, digital zoom is a fast operation and it has a very flexible center of expansion.

display arbitration The process that decides which camera shows the best view of the target object.

F

fixate The process of panning and tilting a camera so as to keep the target object in a camera's view.

H

hybrid zoom A mixture of digital zoom and optical zoom. It takes advantage of the speed of digital zoom's change of magnification and its center of expansion flexibility, with optical zoom's increased detail.

K

Kalman filter A filter which combines inaccurate data points, *e.g.* a model and measurements, in an optimal fashion to estimate the state of a system.

M

magnification How much zoom happens; the ratio of the height of the image compared to the height of the object. Increasing the magnification makes the object appear to grow. Magnification is generally slow for optical zoom while it is nearly instantaneous for digital zoom.

N

notation

- Arrays are shown in bold lowercase (*e.g.* **a**, **b**, **c**).
- Matrices are shown in capital teletype font (*e.g.* **A**, **B**, **C**).
- Subscript $_{k+1|k}$ indicates something calculated about time $k + 1$ given information at time k .
- Images have an apostrophe, *e.g.* the image of point P is shown as P' , and the height of an object, h , has an image of height h' .
- A line between two points, A and B , is represented as AB with length $\overline{AB}$.
- A dot represents derivative, *e.g.* position x has velocity $\dot{x}$ and acceleration $\ddot{x}$.

O

optical axis The line perpendicular to the image plane and passing through the pinhole of a camera.

optical zoom When the size of an image is increased using a lens. The target object's detail increases with magnification. The act of changing the magnification is a slow operation, as it relies on mechanical time constants. While the center of

expansion for optical zoom can change slightly as the lens properties change, the center of expansion is considered fixed when compared to digital zoom.

P

pan Rotating the camera horizontally, along the vertical axis, whereas all objects in the captured image appear to move either left or right.

S

stability Specifically in this work, the ability of the system to maintain fixation. A stable system can maintain fixation, while an unstable system cannot.

T

target object The object in the real world that the user wishes to capture in an image using the system.

tilt Rotating the camera vertically, generally along the horizontal axis orthogonal to the optical axis, whereas all objects in the captured image appear to move either left or right.

track To determine both the position and the size of the target object's image in a series of images.

Chapter 1

Introduction

1.1 Active vision

In the early 1960s, researchers viewed computer vision as a relatively simple problem—if humans and multitudes of other organisms can so effortlessly see, then how difficult can it be to design a man-made system with similar attributes? The perception was that it would be mere decades before we were able to surpass the capabilities of a natural vision system—but nearly half a century later, research has indicated that human vision is considerably more complex than imagined.

Not to say that advances have not been made. It is fair to say that the human eye is similar in function to a digital camera, inasmuch as they both capture an image for processing by either the brain or a computer. It would diminish the significance of this work to say the sensing problem has been solved, however it is the interpretation of the captured image—the task done by the brain—that continues to frustrate researchers. If and when we figure out the perception problem, potential applications include an automated cameramen for sporting events, or—for these days of heightened security awareness—a system that can recognize known terrorists in a crowd, or an easily searchable library of images similar to the text-search capabilities of the modern day Google. The possibilities are seemingly limitless.

A particular application of interest is the ability to recognize and interpret sign language, specifically American Sign Language (ASL). Used as the primary mode of communication for millions worldwide, ASL is a language that communicates not only words but also the full gamut of human emotions, similarly to how inflexions in the spoken word can convey joy, sarcasm, or indifference. Therefore, to understand ASL as a human would, a computer must look at the hands, face, and body language of the signer.

The starting point of this daunting task is to interpret a single hand in real-time. Preceding works present ways to track hands in real-time, such as Clark in [6], and to recognize basic signs, such as Rupe in [22]. The missing bridge between the works is that the captured image must be conditioned appropriately for the recognition algorithm to be effective; object recognition algorithms often have strict scale and position requirements where if the requirements are not met, the performance rapidly degrades to failure. For example, the work of Yang *et al.* in [34] requires a $60 \text{ pixels} \times 60 \text{ pixels}$ image while the work of Heisele *et al.* in [17] requires $58 \text{ pixels} \times 58 \text{ pixels}$. Other algorithms are more resistant to varying scale and position, but still have a preferred size, such as Rupe's work which performs well between $50 \text{ pixels} \times 50 \text{ pixels}$ and $250 \text{ pixels} \times 250 \text{ pixels}$, but works best with smaller images in that range.

The goal of this work is to develop a system that can reliably capture images of a specific size and position determined by a higher-level process (e.g. an object recognition algorithm).

1.2 Thesis outline

Chapter 2 discusses all relevant research done preceding this work, which includes a description of the original man-made imaging system, *i.e.* the pinhole camera, the understanding of which directly leads into how a lens improves image sensing. A benefit of lenses is that they facilitate optical zoom, the control of which is presented by a single-camera

zooming algorithm that aims to maximize focal length while guaranteeing that object fixation will not be lost.

In the event of lost fixation—at least when there is no plan to reacquire fixation, such as in this work—an active vision system becomes unstable; observing a scene, or a specific object in a scene, is the *raison d'être* of the system, therefore if the object is out of view, the system can do nothing. **Chapter 3** proposes that by adding a second camera with fixed focal length, stability can be independent of the zoom action.

With stability no longer a concern, zoom only changes the size and resolution of the captured image. However, slow zoom response continues to prevent the ability to produce scale invariant images for all but slow-moving objects. In an effort to speed up zoom action without changing the hardware, **Chapter 4** uses digital zoom to remove mechanical time constants from zoom, which splits the zoom into two pieces: optical and digital. With digital zoom guaranteeing scale invariance, optical zoom only affects resolution.

Finally, **Chapter 5** concludes the work by listing contributions and recommending future work.

To ensure completeness, Appendix A discusses topics that otherwise do not flow with the rest of the text.

Chapter 2

Background Theory and Related Work

2.1 An introduction to cameras

In the 15th century, Brunelleschi published the first paper describing his understanding of image formation. The pinhole projection model was based on his experience with the original man-made imaging system, the pinhole camera, which is still widely used today among photographers. A pinhole camera is a favorite experiment of early optics courses, and can be easily constructed by poking a small hole in one side of an otherwise enclosed box, then replacing the opposite side of the box with a translucent material. As seen in Figure 2.1, light rays then pass through the pinhole to create an inverted image on the

Figure 2.1: Pinhole camera [12]

Figure 2.2: Object size and object distance affect image size [12]

translucent side. Various scene and camera attributes determine the specific characteristics of that image, including the size, brightness, and orientation.

To demonstrate size, consider Figure 2.2 where object B is twice the height of both A and C , yet the height of image B' is identical to C' and twice that of A' . The size of an object does not directly affect image creation, but rather it affects the angle of the light passing through the pinhole. Notice that object C is half the distance from pinhole O as A and B . Given that it is a geometry problem, the distance is important; the relative size of images depends not only upon size of an object, but also on its distance from the pinhole.

A projection—which may or may not consider this trait—connects the 3-D world with a 2-D image, enabling estimates of the world to be made by measuring an image. Dropping a dimension results in an inevitable loss of information, yet the choice of projection determines how much information will be lost; every projection considers a unique subset of scene and camera attributes to understand the formation of an image.

Figure 2.3: Perspective Projection [12]

2.1.1 Perspective projection

All aspects of the pinhole model are used for the perspective projection, which for convenience is placed in coordinate system (O, i, j, k) —as seen in Figure 2.3—where O is the pinhole; P is a point with image P' ; the image plane Π' lies at distance f' from O , and is orthogonal to k ; C' is the image center where the optical axis intersects Π' .

Point P is mapped to P' by

$$x' = f' \frac{x}{z} \quad (2.1)$$

$$y' = f' \frac{y}{z} \quad (2.2)$$

$$m = \frac{f'}{z} \quad (2.3)$$

where m is the magnification and $z' = f'$ since P' is on Π' .

While accurate, the perspective projection lacks practicality as its computationally intensive nature—exacerbated by the need to know the unique depth of every point—hurts real-time performance as well as potentially causing a problem to be ill-conditioned [30].

Figure 2.4: Weak perspective projection [12]

2.1.2 Weak perspective projection

Some approximations of the perspective projection have shown to be more practical. For example the weak perspective projection considers a group of points to all lie on the fronto-parallel plane, with depth z_0 as denoted in Figure 2.4 as Π_0 . Effectively, perspective effects within a single object are ignored such that point P is mapped to P' by

$$x' = -mx \quad (2.4)$$

$$y' = -my \quad (2.5)$$

$$m = -\frac{f'}{z_0} \quad (2.6)$$

where physical constraints force z_0 to be negative, making the magnification m positive.

The weak perspective projection is classified as an affine projection along with the paraperspective projection, the orthographic projection, and others. Each of these projections may suffice in modeling a camera depending on the requirements of a specific application. Plenty of techniques in computer vision take advantage of affine projections, *e.g.* [15, 23, 30, 12].

(a) Pinhole camera

(b) With lens

Figure 2.5: Lenses allow more light to enter the pinhole [12]

Figure 2.6: Thin lens model

2.1.3 Thin lens model

The projections presented above all assume an ideal pinhole camera, which allows only a single ray of light to reach the image plane. Real pinholes have finite size, such that smaller pinholes create faint images and larger pinholes create blurry images. As shown in Figure 2.5, lenses allow more light to pass through a small pinhole, producing a bright and sharp image.

A lens refracts, or bends, light such that it is redirected through a pinhole, or focal point F . The lens's shape, thickness, and material determines the nature of refraction. With the thin lens model in Figure 2.6, which assumes the thickness of the lens is infinitesimally small, z' is not necessarily equivalent to f' but rather

$$\frac{1}{f} = \frac{1}{z'} - \frac{1}{z} \quad (2.7)$$

where $f = f'$ is the focal length. The magnification is defined as

$$m = \frac{z'}{z} = \frac{h'}{h} \quad (2.8)$$

where h is the height of an object and h' is the height of its image. For a zoom lens, the focal length varies, allowing the magnification to change.

A real zoom lens is generally made up of dozens of individual moving lenses. Depending on the quality of the lens and the application for which the lens is used, the thin lens model can still be applied [30, 33].

For more information about projections and lenses, see [12, 27].

2.1.4 Calibration

To measure an image it is necessary not only to understand the projection, but also to know specific attributes of the camera and lens. Some intrinsic parameters include focal length, skew, principal point, distortion. Extrinsic parameters include rotation and translation.

The acquisition of these attributes is called *calibration* [24, 12]. Depending on the complexity of the system as well as the attributes necessary, calibration can be costly. When zooming, different methods of tracking are used, as shown in [15, 11, 28, 29, 16, 30], all of which require calibration of the zoom lens, a process more costly than constant focal length camera calibration [21].

For some applications, it is possible to consider the effect of certain parameters as negligible. Such assumptions increase the measurement error, but a closed-loop control system can reduce that error. When fixating it is essential to approximate a pixel to camera angle near the center of the image. To acquire this parameter, measurements need to be conducted as shown in Figure 2.7. Rotational calibration can be done *a priori* and therefore does not affect real-time performance.

Similarly, zooming requires knowledge of focal length. If zoom odometry is not available, then modeling of the zoom lens must be done in real-time, which hurts performance.

Figure 2.7: Rotational calibration

If a camera provides real-time odometry, then performance is unaffected. Since off-the-shelf cameras with zoom odometry are available, it is possible to completely model the camera in a manner that is favorable to real-time systems [30].

2.2 Tracking

Camera models explain image creation, but an understanding of the captured image is integral to perception. By following an object through a sequence of images, tracking algorithms can use perception to improve sensing, allowing an object to be fixated upon by a camera. Different trackers can use rigid models [14], optical flow [27], support vector classifiers [2, 22], background subtraction [5], *etc.* to follow objects. This work is concerned more with real-time performance than the accuracy of a particular tracker, so a fast algorithm is preferred. One such method is color segmentation.

Figure 2.8: Image used to train color segmentation tracker

2.2.1 Classification and clustering via color segmentation

This algorithm denotes a range of colors as being part of the target object, or foreground, while every other color is part of the background. Different color spaces of interest are RGB, rgb, and YUV. In this work, the nonlinear color space hue-saturation-value (HSV) is used, sometimes referred to as hue-saturation-intensity (HSI) or hue-saturation-lightness (HSL). Frame grabbers commonly acquire images in the linear RGB space, so an example of a conversion to HSV from [18] is

$$\begin{aligned}
 V &= \max(R, G, B) \\
 S &= \begin{cases} (V - \min(R, G, B)) * 255/V, & \text{if } V \neq 0 \\ 0, & \text{o.w.} \end{cases} \\
 H &= \begin{cases} (G - B) * 60/S, & \text{if } V = R \\ 180 + (B - R) * 60/S, & \text{if } V = G \\ 240 + (R - G) * 60/S, & \text{if } V = B \end{cases} \\
 &\text{if } H < 0 \text{ then } H = H + 360
 \end{aligned}$$

After initial segmentation, the morphological operations of dilation and erosion improve the contiguousness of the foreground—known as clustering. A connected components algorithm then labels each of the separate foreground regions, with the largest of those regions denoted as the target object.

Figure 2.8 shows the image used to train the color segmentation tracker for red objects. As a sidenote, future chapters use this tracker to test ideas presented in this work. While this tracking algorithm is rather simplistic, it does provide a robust and easy-to-maintain algorithm, that is proven in real-time systems; it serves as a proof-of-concept.

2.3 Focal length selection for optical zoom

Tracking information is essential in determining the optimal focal length that maximizes resolution while maintaining fixation because focal length—or more appropriately zoom—acts as a measurement gain, amplifying both resolution and tracking error. Multiple sources present ways to accomplish that, including Tordoff in [30] and Denzler *et al.* in [10].

Sznaier *et al.* propose a different method altogether by using robust control to make an active vision system resistant to changes in focal length. Their method is limited in that wider ranges of focal length degrade performance [26, 25].

The Tordoff method, because it is proven in a real-time environment, is presented below.

2.3.1 A steady-state Kalman filter

From a practical point of view, a Kalman filter is a recursive algorithm that makes estimates and predictions about the state of a system based upon noisy data from multiple sources, such as from a model and measurements. The Kalman filter combines, in an optimal fashion, the data points to estimate the current state and predict future states [1, 20].

The filter’s primary objective is to maintain fixation by controlling the pan angle, tilt angle, and focal length of a camera. The prediction model assumes a constant velocity target, with unmodelled acceleration appearing as noise. The state of the target object is

based upon its angle relative to the optical axis, with state vector $\mathbf{p} = (\phi, \dot{\phi})^\top$ where ϕ is the pan (or tilt) angle of the camera needed to fixate directly upon the target object. For simplicity, the equations only account for a single dimension (*i.e.* pan), but adding extra orthogonal dimensions (*e.g.* tilt) is trivial. The linear model is:

$$\mathbf{p}_{k+1} = \mathbb{F}_k \mathbf{p}_k + \mathbf{u}_k + \mathbf{q}_k \quad (2.9)$$

where subscript k denotes the discrete frame; $\mathbf{u}_k = (-\theta_k, 0)^\top$ is the control input where θ_k is the change in camera angle between time $k - 1$ and k ; and $\mathbf{q}$ is the process noise drawn from a zero mean Gaussian noise sequence with covariance $E[\mathbf{q}_k \mathbf{q}_k^\top] = \mathbb{Q}_k$. For constant velocity motion, the state transition matrix $\mathbb{F}_k$ is,

$$\mathbb{F}_k = \mathbb{F} = \begin{pmatrix} 1 & \Delta T \\ 0 & 1 \end{pmatrix} \quad (2.10)$$

where ΔT is the sampling time, such that $\Delta T = t_k - t_{k-1}$. To complement the model, measurements of an image projected as a 3D to 2D mapping such that

$$\mathbf{m}_k = \mathbb{H}_k \mathbf{p}_k + \mathbf{r}_k \quad (2.11)$$

where $\mathbf{m}$ is the measurement; $\mathbf{r}_k$ is the measurement noise drawn from a zero mean Gaussian noise sequence with covariance $E[\mathbf{r}_k \mathbf{r}_k^\top] = \mathbb{R}_k$; and $\mathbb{H}_k$ is given by

$$\mathbb{H}_k = \mathbb{H} = \begin{pmatrix} 1 & 0 \end{pmatrix} \quad (2.12)$$

where $\mathbb{H}_1 = 1$ exhibits fixed focal length of unity—dynamic focal length would create a time-varying filter. The measurement is then an x -coordinate in a unit width image ($-\frac{1}{2} < x < \frac{1}{2}$).

There are three stages in the Kalman filter: prediction, measurement, and correction. **Prediction** estimates the values and covariances of the state (Equations 2.13 and 2.14) and measurement (Equations 2.15 and 2.16) given *a priori* knowledge. The notation $k + 1|k$

indicates an estimation of the state at time $k + 1$ given all the information available at time k .

$$\hat{\mathbf{p}}_{k+1|k} = \mathbf{F}_k \hat{\mathbf{p}}_{k|k} + \mathbf{u}_k \quad (2.13)$$

$$\hat{\mathbf{P}}_{k+1|k} = \mathbf{F}_k \hat{\mathbf{P}}_{k|k} \mathbf{F}_k^\top + \mathbf{Q}_k \quad (2.14)$$

$$\hat{\mathbf{m}}_{k+1|k} = \mathbf{H}_{k+1} \hat{\mathbf{p}}_{k+1|k} \quad (2.15)$$

$$\hat{\mathbf{M}}_{k+1} = \mathbf{H}_{k+1} \hat{\mathbf{P}}_{k+1|k} \mathbf{H}_{k+1}^\top + \mathbf{R}_{k+1} \quad (2.16)$$

In the **measurement** stage, the expected measurement is compared to the actual measurement, giving the innovation $\boldsymbol{\nu}_{k+1}$ defined as

$$\boldsymbol{\nu}_{k+1} = \mathbf{m}_{k+1} - \hat{\mathbf{m}}_{k+1|k} \quad (2.17)$$

The **correction** stage then recalculates the state based on *a posteriori* knowledge of the measurement as follows:

$$\mathbf{W}_{k+1} = \hat{\mathbf{P}}_{k+1|k} \mathbf{H}_{k+1}^\top \hat{\mathbf{M}}_{k+1}^{-1} \quad (2.18)$$

$$\hat{\mathbf{p}}_{k+1|k+1} = \hat{\mathbf{p}}_{k+1|k} + \mathbf{W}_{k+1} \boldsymbol{\nu}_{k+1} \quad (2.19)$$

$$\hat{\mathbf{P}}_{k+1|k+1} = \hat{\mathbf{P}}_{k+1|k} - \mathbf{W}_{k+1} \hat{\mathbf{M}}_{k+1} \mathbf{W}_{k+1}^\top \quad (2.20)$$

Initial values for $\hat{\mathbf{p}}_0$, $\hat{\mathbf{P}}_0$, $\mathbf{Q}$, and $\mathbf{R}$ must be provided. The derivation of the filter update equations is available in the work of Bar-Shalom *et al.* [3].

2.3.2 Zoom invariant Kalman filter

Adding zoom to the filter does not change the model in Equation 2.9, but it does affect the measurement in Equation 2.11, such that

$$\mathbf{H}_k = \begin{pmatrix} f_k & 0 \end{pmatrix} \quad (2.21)$$

where f_k is the camera's focal length at time k . With non-constant $\mathbf{H}$, the filter becomes a linear time-varying filter.

To simplify the time-varying filter, Tordoff presented a method to make the filter behave identically to the steady-state filter by either scaling measurement noise with focal length

$$\mathbf{R}_k = f_k^2 \mathbf{R} \quad (2.22)$$

or scaling process noise with the inverse of focal length, $\alpha_k = 1/f_k$

$$\mathbf{Q}_k = \alpha_k^2 \mathbf{Q} \quad (2.23)$$

By using either equation 2.22 or 2.23, the filter is unaffected by changing focal length, but using both equations actually improves fixation error—as shown by simulation.

Simulations

For proper comparisons, each of the following trials—originally presented in [30]—undergo the same motion: an object begins at position $\phi = -60^\circ$ with velocity $\dot{\phi} = 30^\circ s^{-1}$, and continues until $\phi = 30^\circ$ whereupon constant acceleration is applied until $\dot{\phi} = -30^\circ s^{-1}$; the object then continues until $\phi = -60^\circ$ —Figure 2.9(a) and Figure 2.9(b) illustrate that motion. Figure 2.9(c) shows the focal length for trials involving variable zoom.

The initial conditions for the trials are $\mathbf{p} = (m_1, m_1 - m_0)^\top$, $f = 1$, $\mathbf{r} = .02$ (approximately 3.2 pixels in a 160×120 image), and $\mathbf{Q}$ as derived in Appendix A.1 is

$$\mathbf{Q} = 10^{-6} \times \begin{pmatrix} \frac{\Delta T^3}{3} & \frac{\Delta T^2}{2} \\ \frac{\Delta T^2}{2} & \Delta T \end{pmatrix} \quad (2.24)$$

where $\Delta T = \frac{1}{30} s$. The pan/tilt platform is assumed ideal in that $\theta_{verg,k+1} = \hat{\phi}_{k+1|k}$. Noise is added to the measurements with standard deviation of 0.02.

The trial with fixed focal length, Figure 2.10, serves as a baseline; a zoom-invariant filter should match this response.

By varying focal length, but keeping process and measurement noises constant, it is evident in Figure 2.11 that zoom adversely affects prediction accuracy. In addition, fixation error increases with zoom in, but decreases with zoom out.

Figure 2.9: Object motion used in figures 2.10-2.14 [30]

Figure 2.10: Tracking with fixed focal length [30]

Figure 2.11: Tracking with variable focal length [30]

Figure 2.12: Tracking with variable focal length, with $R \propto f^2$ [30]

Figure 2.13: Tracking with variable focal length, with $Q \propto \frac{1}{f^2}$ [30]

Figure 2.14: Tracking with variable focal length, with $R \propto f^2$ and $Q \propto \frac{1}{f^2}$ [30]

By scaling measurement noise with focal length, Figure 2.12 shows prediction identical to the baseline—a zoom-invariant filter. The fixation error is still affected by zoom level. Figure 2.13 presents another zoom-invariant filter while scaling process noise with the inverse of focal length.

Figure 2.14 shows that estimation actually improves when using both Equation 2.22 and 2.23. This filter is the best fitted for changing focal length; performance of the filter is not zoom invariant but rather it is *improved* by the presence of zoom.

For more details about the simulations, see [30].

2.3.3 Controlling focal length

The preceding simulations used predetermined focal length, proving that changing focal length affects fixation error. This section looks to maximize zoom while keeping fixation error bounded; the amount of fixation error is unimportant as long as the target object remains in the image.

A critical piece of zoom control is the innovation calculated from the Kalman filter, which represents the error between what is expected in the constant velocity model and what is observed. This error encapsulates target object acceleration, and other noise present in the system. The desired focal length is inversely proportional to the variance of the innovation.

From [30]:

In the context of Kalman filtering, [bounded error] can be interpreted as specifying a confidence bound on the innovation

$$p(|\nu| < \psi) \geq \zeta \quad (2.25)$$

where ψ is the boundary location, and ζ the confidence that is required. If the innovation is 1D, as in the previous simulations, then for a matched filter ν forms a Gaussian white noise sequence, allowing the ideal variance of the innovation to be specified. For instance, if $\zeta = 99.9999\%$ (the “one in a million”

confidence interval) then the ideal variance is $\text{var}[\nu] \approx \psi^2/24$. Since ν scales with focal length, we wish to zoom out if the variance of ν is too high, and the control law becomes

$$f_{k+1}^2 \leq \frac{\psi^2}{\text{var}[\nu_k]} \quad (2.26)$$

The case of a 2D innovation is similar, as ν again scales with zoom. For a matched filter ν is sampled from a 2D Gaussian, which will generally not be independent in the image axes. . . The requirement is then that the maximal variation in any direction is within the bound ψ . For the “one in a million” bound, the control law is

$$f_{k+1}^2 = \frac{\psi^2}{24 \|\text{covar}[\nu_k]\|_2} \quad (2.27)$$

where the 2-norm of the covariance gives the maximum variation (*i.e.* the largest eigenvalue of the covariance matrix). To enforce [maximum resolution,] the inequality is replaced by equality, and zoom is controlled to always make the covariance touch this boundary.

He continues:

In order to fulfil the bounding conditions, an estimate of the innovation (co-)variance is required. For changing target dynamics, this variance will change over time, and so the statistics of the most recent errors are required. A simple approach is to maintain an exponentially weighted sequential estimate of the covariance

$$\text{covar}[\bar{\nu}_k] \approx \gamma (\nu_k \nu_k^\top) + (1 - \gamma) \text{covar}(\bar{\nu}_{k-1}) \quad (2.28)$$

where γ is the “forgetting factor” in the range $0 < \gamma < 1$, and varying γ as $1/k$ would give unweighted estimates. The choice of forgetting factor can be particularly important, as a long memory ($\gamma \ll 1$) aids smoothing but takes longer to react to changes, but a short memory ($\gamma \gg 0$) can give noisy estimates. . .

(a) $\gamma = 0.02$

(b) $\gamma = 0.10$

(c) $\gamma = 0.50$

Figure 2.15: The effect of varying the “forgetting factor” γ [30]

In practice, there are some subtleties to such a scheme when zoom is changing. First as ν_k is scaled by f , the variance calculation must take into account changes of zoom. The zoom-normalized variance is therefore calculated:

$$\text{covar} [\bar{\nu}_k] \approx \gamma \frac{(\nu_k \nu_k^\top)}{f_k^2} + (1 - \gamma) \text{covar} [\bar{\nu}_{k-1}] \quad (2.29)$$

Practically, these equations cause the camera to zoom-in when the velocity is constant and zoom-out whenever there is any acceleration. A larger γ allows for quick adjustments in zoom level, while a small γ allows for slower adjustments, see Figure 2.15.

Zooming in can cause instability as it increases the possibility of the target object leaving the image, while zooming out only costs detail. Calculating two covariances, one with a large γ and another with a small γ , then using the value with the small focal length creates a conservative zoom-in and a faster zoom-out. This method is safe with respect to bounding the fixation error, but at the expense of resolution.

2.4 Center of expansion

The focal length determines how much zoom, but it does not determine where to zoom. Optical zoom lenses generally zoom around the center of the image, called the *center of expansion*, defined by Willson in [33]:

Given two images taken at different magnifications, exactly one position in the scene in both images will remain in the same place on the image plane. This position is called the “center of expansion” between the two images. More precisely, given two images I_1 and I_2 taken at two magnifications m_1 and m_2 , and given n reference points $P_1 \dots P_n$ in image I_1 and the corresponding points $Q_1 \dots Q_n$ in image I_2 , the center of expansion C satisfies the constraint

$$(C - P_i) = k (C - Q_i) \quad \forall i = 1 \dots n \quad (2.30)$$

where

$$k = \frac{m_1}{m_2} \quad (2.31)$$

(a) Zoomed out view

(b) Zoomed with top-center center of expansion (c) Zoomed with bottom-left center of expansion

Figure 2.16: Example of center of expansion

Willson describes center of expansion in the context of calibrating zoom lenses, but it applies for both optical and digital zoom. Figure 2.16 shows an example of 2.8X zoom magnification with different centers of expansion. The magnification determines the size of the image, while the center of expansion determines what is shown in the image.

For optical zooming, the center of expansion is usually the same as—or very close to—the principal point of the lens, which is the intersection of the optical axis with the image plane. The center of expansion can change slightly as other lens parameters change, but can generally be considered fixed. A fixed center of expansion is difficult when trying to zoom on a specific object; if that object is not on the principal point, then the object will eventually fall off the edge of the image when zooming in. The only way to move the center of expansion onto the target object is to either a) move the target object or b) move the image axis and image plane, *e.g.* pan and/or tilt the camera. Assuming that the former is not possible, the only choice is to move the camera. Therefore, the goal of scale invariance is not only limited by zoom response but also pan and tilt response.

With digital zoom or advanced zoom lenses [32], the center of expansion is more flexible as it can be instantaneously placed anywhere on the image plane.

2.5 Summary

For image perception, it is important to understand how images are captured. While not well-suited for real-time applications, the perspective projection is an accurate model of a camera. The less precise weak perspective projection ignores most perspective effects, making it less computationally intensive and therefore appropriate for real-time applications. Combined with a lens model and a tracking algorithm, projections describe the sensing aspect of active vision.

To improve sensing, real-time optical zoom—enabled by changing focal length in a lens—not only can boost image resolution, but also can eliminate fixation error. However, a tradeoff is necessary as one cannot simultaneously have minimal fixation error and maximal

resolution. This compromise was addressed by maximizing resolution while bounding fixation error.

The center of expansion for digital zoom is free, while for optical zoom lenses it is fixed.

Chapter 3

Dual Camera Zoom

3.1 Introduction

The presence of zoom can improve the perceptibility of an acquired image by either magnifying detail or broadening the field of view. However, since higher focal lengths confine the periphery, both improvements cannot occur simultaneously.

Tordoff elaborates in [30] saying “consider a camera operator viewing a stationary object (it might be a golf ball on the fairway, or a gnu on the veld). While stationary, the operator’s instinct is to zoom in. However, as soon as the object starts to move, the camera-man will react both by attempting to track and by zooming out. As tracking is restored, say at constant angular velocity, the operator may have sufficient confidence to zoom in again, but if subsequently the tracked object moves unexpectedly he will surely once more zoom out. It appears that the camera operator is reducing tracking error to an acceptable distance in the image, where ‘acceptable’ means better than half the image dimension—at worst he wishes to retain sight of the object on the image plane.”

Now consider that the camera operator does indeed lose sight of the object. Recognizing that, he pulls his eye away from the camera, finds the object, and then adjusts the camera accordingly. In this case, the naked eye serves as a fixed focal length, panoramic camera.

This chapter expands on Tordoff’s work by introducing a second camera that serves as the naked eye, allowing a more aggressive zoom without risking overall system instability.

Similar methods are presented in the works of Greiffenhangen *et al.* in [13] and Huang *et al.* in [19]. They differ from this work in that their panoramic camera serves as a stationary overseer—it is mounted on the ceiling. In this work, both cameras rotate from nearly the same viewpoint.

This chapter explains how to detect when fixation is lost, then subsequently how to recover. A real-time system which implements the algorithm is presented along with experiments that validate the method.

3.2 Recovering fixation

3.2.1 Camera view correspondence

Communication between two rotating cameras requires *camera view correspondence*. Consider Figure 3.1 where Z and P , included as either labels or subscripts, represent the zooming and panoramic cameras; $\mathbf{k}$ is on the optical axis of camera Z at an arbitrary but constant time t_0 ; image centers C'_Z and C'_P are situated such that line $C'_Z C'_P$ is parallel to $\mathbf{i}$ with length d ; image planes Π'_Z and Π'_P are both orthogonal to $\mathbf{k}$ at time t_0 . The center of the target object, T , is related to the image centers such that

$$(\overline{C'_P T})_i = (\overline{C'_Z T})_i + d \quad (3.1)$$

$$(\overline{C'_P T})_j = (\overline{C'_Z T})_j \quad (3.2)$$

$$(\overline{C'_P T})_k = (\overline{C'_Z T})_k \quad (3.3)$$

where subscripts $\mathbf{k}$ and $\mathbf{i}$ indicate the components of the line along the appropriate axis, *e.g.* the pythagorean theorem reads as

$$\overline{C'_P T} = \sqrt{(\overline{C'_P T})_i^2 + (\overline{C'_P T})_j^2 + (\overline{C'_P T})_k^2} \quad (3.4)$$

Figure 3.1: Camera correspondence

The angle of line $C'_Z T$ on the plane with basis of i and j is represented by pan angle θ_Z . Similarly, the angle of line $C'_P T$ on the same plane is θ_P . The same lines have tilt angles ϕ_Z and ϕ_P on the plane with basis of i and k .

Since C'_Z and C'_P differ only by a translation in the i direction, then the tilt angles are identical

$$\phi_Z = \phi_P \quad (3.5)$$

leading to the simplified view shown in Figure 3.2 where

$$x = (\overline{C'_Z T})_i = (\overline{C'_P T})_i - d \quad (3.6)$$

$$z_0 = (\overline{C'_Z T})_k = (\overline{C'_P T})_k \quad (3.7)$$

where z_0 is the distance to the fronto-parallel plane. The pan angles are related by

$$\theta_Z(\theta_P, d, z_0) = \tan^{-1} \left[\left(\cot(\theta_P) - \frac{d}{z_0} \right)^{-1} \right] \quad (3.8)$$

Figure 3.2: Camera correspondence overhead view

3.2.2 Assisted control

When fixating with a single camera, the desired camera angles— θ and ϕ —place T' at C' . When T' is on Π' , then a tracker can be used to calculate the angles, called *autonomous control*. However if T' is off the edge of Π' , then *assisted control* is necessary, which requires a second camera.

Consider frame k where T' is on Π' for both cameras, then $z_{0,k}$ is the distance of the image planes to the fronto-parallel plane

$$z_{0,k} = \frac{d}{\cot(\theta_{P,k}) - \cot(\theta_{Z,k})} \quad (3.9)$$

Now say for some future frame $n > k$ that T'_Z is off image plane Π'_Z , but T'_P remains on Π_P . Then θ_P and ϕ_P can be computed using assisted control. Equation 3.5 gives ϕ_Z , and θ_z can be found using Equation 3.8 if z_n is known. Using the basic lens equation, originally presented as Equation 2.8,

$$\frac{z'}{z} = \frac{h'}{h} \quad (3.10)$$

where h is the height of the object, then for frame n and k

$$z_n = \frac{h_n}{h'_n} z'_n \quad (3.11)$$

$$z_k = \frac{h_k}{h'_k} z'_k \quad (3.12)$$

By assuming image distance z' and object height h do not change over time, *i.e.*

$$z'_k = z'_n \quad (3.13)$$

$$h_k = h_n \quad (3.14)$$

then

$$\frac{z_n}{z_k} = \frac{h'_k}{h'_n} \quad (3.15)$$

$$z_n = z_k \frac{h'_k}{h'_n} \quad (3.16)$$

In addition, to account for object rotations about the optical axis, one may also consider measured width w' , combining it with height to give measured area $A' = h'w'$. Now

$$z_n = z_k \sqrt{\frac{A'_k}{A'_n}} \quad (3.17)$$

which with Equation 3.8 gives the desired pan angle.

Sources of error

The performance of assisted control depends upon precision of the angle, position, and height measurements. Without accurate calibration of the cameras, camera odometry may be used for measurements of θ and ϕ . For that case, it is recommended that frame k correspond to when the points T'_Z and T'_P are as close to their respective image centers C'_Z and C'_P as possible, to minimize the effect of inaccurate calibration.

In addition, Equations 3.13 and 3.14 assume that the object height and image distance remain constant. The truthfulness of that statement must be considered when implementing assisted control.

3.2.3 Control arbitration

Like most vision tasks, switching between assisted control and autonomous control is an action naturally suited to humans, as the Cartesian Theater holds assumptions about the scene [9]. However, given the state of modern computer perception, particularly real-time perception, a single-camera has no notion of how an object *should* look; deciphering between target deformities, rotations, and fixation losses is not obvious.

In a dual-camera system, a comparison can be made between the cameras' images—if measurements do not make sense, then fixation must be lost. An assumption that camera P always has the object in view is made for simplicity; camera P is more stable, so if it loses fixation then the system is unstable.

The target object is considered in view of the zooming camera when the following conditions hold true

$$h'_P - \Delta \leq h'_Z \frac{f_P}{f_Z} \leq h'_P + \Delta \quad (3.18)$$

$$w'_P - \Delta \leq w'_Z \frac{f_P}{f_Z} \leq w'_P + \Delta \quad (3.19)$$

where Δ is the uncertainty of the tracker's results; w' is the measured width; h' is the measured length; f is the focal length; and subscripts Z and P represent the zooming and panoramic cameras, respectively. If both of these hold true, then the *control arbitrator* may decide to allow camera Z to resume autonomous control.

The transition from autonomous control to assisted control is seamless, but the arbitrator must not resume autonomous control too soon as the zooming camera's filter will have gone $n - k$ frames without receiving a measurement. A safety delay, t_{safety} , to allow camera Z 's filter to receive measurements is recommended such that

$$t_{\text{safety}} = \min(\alpha t_{\text{assisted}}, t_{\text{max}}) \quad (3.20)$$

Figure 3.3: Sony EVI-D100

where $t_{\text{assisted}} \equiv n - k$ is how long assisted control was used, α is a weight appropriate for the filter, and t_{max} is the maximum possible amount of safety time needed to regain stability in the event of lengthy object loss.

3.3 The experimental system

To test the effectiveness of assisted control, a system was constructed using a pair of Sony EVI-D100 cameras, shown in Figure 3.3.

The pan angle, tilt angle, and focal length can be sampled from the camera every frame, at 30Hz, but control signals can only be sent every other frame, 15 Hz. A large control action for are sent simultaneously for pan/tilt and zoom the camera will fault and ignore commands. Given these limits of the camera, it proved sufficient as a proof-of-concept.

Figure 3.4: Step response for pan and tilt

Communication from the cameras to the controller utilized the Sony VISCA protocol over RS-232 lines, connected to a PC powered by two Pentium Xeon 2.8 GHz with Hyperthreading CPUs. Two Osprey 100/200 were used as a frame grabber, which have the capability of capturing images at 30 fps at 320 x 240 non-interlaced or 640 x 480 interlaced.

3.3.1 Pan and tilt

Figure 3.5 shows the step response for pan and tilt. Pan and tilt have similar plants, which differ on by velocity limits: 300°/s for pan and 100°/s for tilt. Through experimentation, the transfer function was found to be approximately

$$T(s) = \frac{1}{\left(1 + \frac{s}{15}\right)^2} \quad (3.21)$$

which includes neither the velocity limit or the 100ms delay between control signal and response.

Figure 3.5 shows the control loop used for controlling pan or tilt, where the reference signal is the center of the image, while image measurements are calculated by considering

Figure 3.5: Block diagram for pan and tilt

both object motion and camera angle. Calibration parameters convert the image measurements back to physical angles.

3.3.2 Zoom

Since the EVI-D100 features focal length odometry, a zoom lens model is unnecessary. However, the odometry is not ready out-of-the-box as a number between 0 and 16384 inclusive—referred to as the raw reading—represents focal length. The step response of the raw reading is shown in Figure 3.6(a), showing a 150 ms delay to initial response and a settling time of 1.5 s. The response is velocity-limited to $7000 \frac{\text{units}}{\text{s}}$.

Figure 3.6(b) shows the lookup table converting the raw reading to magnification, which was generated by zooming to various readings and measuring the area of an object. The step response of the magnification is shown in Figure 3.6(c). Pseudocode for lookup table acquisition is given in Figure 3.7, where the red square in Figure 3.8 was used as the target object.

Figure 3.6: Zoom reading step response, lookup table, and magnification step response

1.	for $n = 0$ to N
2.	$reading \Rightarrow (\text{max reading} - \text{min reading}) * \frac{n}{N}$
3.	Zoom to $reading$
4.	Wait for zoom to complete
5.	$size[n] \Rightarrow$ average size of target in m images
6.	$magnification[n] \Rightarrow size[n] / size[0]$
7.	end for

Figure 3.7: Pseudocode to create magnification lookup table

The nonlinearities in zoom action—particularly the velocity limit—restricts the effectiveness of linear control theory in improving zoom response.

Figure 3.8: Zoom reading calibration

Figure 3.9: Control arbitration block diagram

3.3.3 System organization

The control of each camera is almost entirely independent, as shown in Figure 3.9, allowing for parallelism to help real-time performance; control arbitration is the only step needing synchronization.

3.4 Experiments on hardware

3.4.1 Pendulum experiment

In the interest of comparison via repeatability, the first experiment is a horizontally swinging pendulum with a red ball as its bob. The parameters used in the test include

$$\begin{aligned}\Delta &= 3 \text{ pixels} \\ \mathbf{r} &= 0.01 \text{ (3 pixels / 320 pixels)} \\ R &= \mathbf{r}^2 = 10^{-4} \\ \gamma &= 0.1 \\ d &= 13 \text{ cm} \\ \Delta T &= 33 \text{ ms} \\ t_{max} &= 1 \text{ s} \\ \alpha &= 1 \\ Q &= 10^{-6} \begin{pmatrix} \frac{\Delta T^3}{3} & \frac{\Delta T^2}{2} \\ \frac{\Delta T^2}{2} & \Delta T \end{pmatrix}\end{aligned}$$

See Appendix A.1 for derivation of Q .

The pendulum is held stationary for the first 1.5 s before being allowed to swing freely for nearly 30 s. Through other experiments not discussed here, the single-camera method was found to be marginally stable with $\psi = 2.0$. With the single-camera method marginally stable, the increased stability of the dual-camera method with $\psi = 4.0$ is evident.

Figure 3.10 shows how the target object is lost and recovered using snapshots from the dual-camera trial. In frame 30, the ball is stationary, indicated by the large magnification of the zooming camera. By frame 43, the ball has been released, showing that the zooming camera quickly lost the ball. In frame 215 the ball has been recovered but leaves the view by frame 231, then again re-enters in frame 235 before being fully in view by frame 239.

All results of the pendulum experiment are contained in Figure 3.11. The zooming camera's fixation is illustrated in Figure 3.11(a), where the expected image height is denoted by the dotted black line, and the measured image height by the solid area. When the expected and measured height do not match, fixation is lost. The camera arbitrator, Figure 3.11(d), enabled the pan angle of the zooming camera to follow the bob, Figure 3.11(c).

Because of the less aggressive zoom, the single camera trial keeps the object in view throughout Figure 3.11(b). The average zoom of the dual-camera method is 1.5X better despite frequently resorting to the panoramic view, as shown in Figure 3.11(e) and Figure 3.11(f). The zoom improvement I is calculated as

$$I = \frac{f_{dual}}{f_{single}} \quad (3.22)$$

where f_{dual} and f_{single} are the focal lengths for the dual- and single-camera methods, respectively.

Figure 3.10: Snapshots of pendulum experiment.

(a) Fixation of dual-camera method

(b) Fixation of single-camera method

(c) Dual-camera pan angles

(d) Decisions of control arbitrator

(e) Side-by-side zoom level comparison

(f) Zoom level improvement. The mean improvement is 1.5

Figure 3.11: Pendulum results

3.4.2 Car experiment

Of course, the primary objective of dual-camera zoom is not to increase zoom level, but rather to increase stability in the presence of zoom. A second experiment was run to further exhibit this trait. One drawback of the pendulum experiment is that if the zooming camera loses fixation, the nature of pendulum motion allows the camera to regain fixation periodically; see Figure 3.11(b) for time 1–3 s.

A radio-controlled car provides sporadic movements that demonstrate the advantages of having a second camera. In fact, the experiment could not be run on the single-camera method without nearly eliminating zoom. This car experiment uses parameters identical to the pendulum experiment, except with $\Delta = 5$ pixels to account for the irregular shape of the car, and $\psi = 4.0$ to keep the zoom aggressive.

Frames 85, 415, and 800 of Figure 3.12 show large magnification in the zooming camera, while frame 260 shows similar zoom levels between the two cameras. Frames 225 and 700 show complete losses of fixation in the zooming camera, which correspond to the fast horizontal movements shown in Figure 3.13(a). Fixation loss is especially evident at $t = 8$ s, $t = 18$ s, and $t = 22$ s in Figure 3.13(c). The system was always able to recover fixation, while maintaining high magnifications as shown in Figure 3.13(d).

Figure 3.12: Snapshots of car experiment.

(a) Dual-camera pan angles

(b) Decisions of control arbitrator

(c) Fixation of dual-camera method

(d) Zooming camera zoom level method

Figure 3.13: Car results

3.5 Summary

The goal of this chapter was not to simply improve stability in the presence of zoom, but rather to make stability independent of zoom. A human camera operator's method of object recovery was used as motivation.

The fixation system needed information about how the target object should look before a determination of quality fixation could be asserted. This information was collected using a second camera. An understanding of the geometry comprising *camera view correspondence* was integral in understanding key concepts such as *assisted control*, which is when one camera is controlled using another camera's sensors. If a camera loses fixation it must switch to assisted control to prevent system instability. When a camera has enough information to control itself, the term *autonomous control* is appropriate. The *control arbitrator* decides which type of control is appropriate.

Experiments showed that adding a second camera with fixed focal length produced a system with stability comparable to a system without zoom. One experiment showed the increase of zoom level by an average of 1.5X over a similar single-camera method.

Chapter 4

Digital Scale Invariance

The long time constants of zoom lenses has proven to be a hindrance not only to stability, but also to scale invariant tracking. A zoom lens must follow object motion to maintain image size, so a slow-moving lens is limited to all but slow-moving objects. An alternative way to change scale is through digital zoom—a concept integral to digital imaging—which magnifies via interpolation and decimation.

While digital zoom is fast, optical zoom’s most valuable characteristic is its ability to adjust image detail and context. The goal of this chapter is to create a hybrid zooming technique which combines the strengths of digital and optical zoom to mask their individual weaknesses.

4.1 Choosing magnification for digital zoom

The exact magnification that provides scale invariance may be obtained in real-time now that zoom changes instantaneously. To preserve image size, it is necessary that the magnification f is proportional to the object distance z_0 . Using the measured area to approximate z_0 can handle translations in any direction, but cannot handle rotations well. This method falls victim to the “spinning coin” problem of unwanted zoom oscillations, as seen in Figure 4.1. Area is smaller when only the edge is visible which causes increased zoom, and vice versa. A solution to this problem is available, but it only applies to planar objects [31].

(a) Spinning coin. [http://www.startcreative.co.uk/images/Coin rotate.jpg](http://www.startcreative.co.uk/images/Coin%20rotate.jpg)

(b) Adjusting focal length with image area [31]

Figure 4.1: Spinning coin problem

An alternative solution to prevent unwanted zoom oscillations is to consider the measured length and width separately. If height alone is used to calculate z_0 , then zoom oscillations do not appear; the width is what oscillates. Of course, this is not a general solution because a different situation could present height as the problem instead of the width.

Therefore, to consider both height and width, the magnification m to produce scale invariance is

$$m_k = \frac{h'_0}{\max(h'_k, w'_k)} \quad (4.1)$$

where h'_0 is the desired image dimension, h'_k is the image height, and w'_k is the image width. All values are for frame at time k . This method creates a bounding square around the object which is resistant to the spinning coin.

4.2 Fixating with digital zoom

Digital images not only allow for rapid zoom, but also for movable image center which can be adjusted to match the tracker's position measurement. Therefore, fixation error can be eliminated. The reduction of fixation error has been studied in earlier chapters, as well as in many works such as [8, 7].

4.3 Noise analysis

4.3.1 Impact of image noise

Image measurements inherently have noise. When measuring the height of an image,

$$h'_{m,k} = h'_k + \mathbf{r}_k \quad (4.2)$$

where $h'_{m,k}$ is the measured height, h'_k is the actual height, and $\mathbf{r}$ is image noise drawn from a zero mean Gaussian noise sequence with covariance $E[\mathbf{r}_k \mathbf{r}_k^T] = \mathbf{R}_k$.

A large signal-to-noise ratio (SNR), *i.e.* $h'_k \gg r_k$, can make noise effects negligible, where

$$\text{SNR} = \frac{h'_k}{r_k} \quad (4.3)$$

However since the SNR varies with measured height, large magnifications are more susceptible to noise. The effect of image noise varies with resolution, which manifests itself as shaky zoom. To decrease this shakiness, magnification can be calculated such that

$$m = \frac{h'_0}{\hat{h}'_k} \quad (4.4)$$

where $\hat{h}'_k$ is the estimated image height. When SNR is large, the measurement can be assumed more accurate than when SNR is small—consequently when measurements are noisy, a model is appropriate.

The height measurement depends not only on noise, but also upon the state of the system such that

$$h'_k = \frac{f_k h_k}{z_k} \quad (4.5)$$

which was derived from Equations 2.8 and 2.7 where f_k is the camera focal length, z_k is the distance to the target object, and $z'_k = f_k$ because $z_k \gg z'_k$.

A filter which does not account for dynamic SNR may only be effective under certain conditions. A Kalman filter as presented in Section 2.3 will be used.

4.3.2 Simulations

To make clear the need for a proper filter, experiments are run using various filters—each of which is fed identical object motion and measurement noise. An object of height 0.35m begins at $z = 0.5$ m with velocity $\dot{z} = 0.5$ m/s away from the camera. Upon reaching $z = 2$ m, constant acceleration is applied until $\dot{z} = -0.5$ m/s. The object continues until reaching $z = .5$ m, where it accelerates once again to $\dot{z} = 0.5$ m/s and remains at that velocity through the end of the simulation. The maximum distance is $z = 2.4$ m and the minimum is $z = 0.1$ m, which corresponds to 4 and 120 pixels, respectively, as shown in

Filter	Error (approx 2-norm)
None	318.2
High image noise covariance	1223.2
Low image noise covariance	186.1
Variable image noise covariance	111.7

Table 4.1: Filter comparison in presence of image noise

Figure 4.2(c). The focal length throughout simulation is $f = 35$ mm. When applicable,

$$\mathbf{Q} = 10^{-6} \times \begin{pmatrix} \frac{\Delta T^3}{3} & \frac{\Delta T^2}{2} \\ \frac{\Delta T^2}{2} & \Delta T \end{pmatrix}$$

The unfiltered run is shown in Figure 4.2(a). Notice that the noise escalates for large z .

Figure 4.3 shows a Kalman filter with $\mathbf{R} = 10^{-4}$. This filter relies substantially on the model, which makes it resistant to the high-noise measurements: a desirable behavior. However, when noise is low, an unwanted delay appears between actual state and estimation.

Figure 4.4 on the other hand has a smaller $\mathbf{R} = 10^{-8}$. It does not rely heavily upon the model for the simulated distances, which leads to less resistance to noise.

A filter which adapts itself to changing conditions is shown in Figure 4.5, where $\mathbf{R}_k = \frac{\mathbf{R}}{h_m'^2}$ and $\mathbf{R} = 10^{-8}$. The result is a filter that is resistant to high noise without adversely affecting the response when noise is low.

The results are organized in Table 4.1. The filter best suited for digital zooming is the one which considers varying measurement noise.

(a) Estimation

(b) Percent error

(c) Distance vs Pixels

Figure 4.2: Unfiltered digital zoom

Figure 4.3: Kalman filter of digital zoom with $R_k = 10^{-4}$.

Figure 4.4: Kalman filter of digital zoom with $R_k = 10^{-8}$.

Figure 4.5: Kalman filter of digital zoom with $R_k = \frac{R}{h_m'^2}$.

4.4 Display arbitration

Similar to control arbitration in Chapter 3, *display arbitration* is needed to determine when the object is in view of the zooming camera. Poorly designed display arbitration will not result in loss of fixation, but rather it will produce a final image that does not contain the object. Therefore, display arbitration parameters Δ , t_{max} , and α may be independent of those used for control arbitration.

In the event of object loss on the zooming camera, the system will switch to the panoramic camera's view. Once the zooming camera regains fixation, the arbiter can return the view to the zooming camera.

In the event that full scene context is preferred over high resolution—as digitally zooming out leaves an undefined region where context would normally be—the arbitration function can be altered appropriately.

Figure 4.6: Digital zoom block diagram

4.5 System Organization

Figure 4.6 illustrates the system setup. Once again, the data path for two cameras are entirely independent with the exception of display arbitration. Please note that control arbitration is omitted from this diagram for simplicity.

4.6 Experiments on hardware

4.6.1 Pendulum experiment

The first experiment combining control and display arbitration is a pendulum swinging towards and away from the cameras to test the speed of zoom—as opposed to the horizontal motion in Section 3.4.1 that tested fixation. Parameters identical to the those in Section 3.4.1 were used, except with $\psi = 4.0$.

Note that the data in the following figures was generated by post-processing the captured video using color segmentation, therefore the effects of the filter may appear as inaccuracies. A better way to acquire the data would have been to use a more robust algorithm, which was not available at the time.

The effectiveness of digital zoom can be seen by comparing object size in the image before and after digital zoom, as shown in Figure 4.7. The solid curve represents image height of the original image, the dotted curve is height for the zoomed image, and the dotted horizontal line at 100 pixels represents the goal. Error is considerably decreased in the digitally zoomed image.

Figure 4.8 compares the fixation error, which exhibits the image re-centering property of digital zoom. Digital zoom appears to entirely eliminate fixation error.

The combination of optical and digital zoom into hybrid zoom is shown in Figure 4.9. The optical zoom is closer to the target than the fixed focal length camera, but digital zoom is able to further reduce error.

Figure 4.7: Scale invariance for panoramic camera

Figure 4.8: Digital fixation for panoramic camera

Figure 4.9: Scale invariance for zooming camera

Figure 4.10 shows this method is able to neutralize the fixation error inherent with large focal lengths.

Selected screenshots of the experiment are shown in Figure 4.11. Digital zoom is always able to fixate on the ball in the panoramic camera. However for the zooming camera, digital zoom cannot improve fixation when the ball is not in the view, particularly in frame 90 and 211. The dark border around the zooming camera's digital images in frames 30, 90, 211, 272, and 548 indicates a digital zoom out, as digital zooming is unable to recover scene context not available in the original image.

Figure 4.10: Digital fixation for zooming camera

Figure 4.11: Display arbitration pendulum experiment

Figure 4.12: Scale invariance for panoramic camera

4.6.2 Car experiment

A second experiment, identical to the one in Section 3.4.2, involves a radio-controlled car moving sporadically. Once again, error in scale invariance, Figure 4.12, is considerably decreased by digital zoom, as is fixation error in Figure 4.13. Hybrid zoom is shown in Figure 4.14 and Figure 4.15.

Figure 4.16 shows selected screenshots. The digitally zoomed images appear identical between the panoramic and zooming cameras, with the exception that the zooming camera's have more detail. Frame 362 highlights this: the key feature of hybrid zoom.

Figure 4.13: Digital fixation of panoramic camera

Figure 4.14: Scale invariance of zooming camera

Figure 4.15: Digital fixation of zooming camera

Figure 4.16: Display arbitration car experiment

4.7 Summary

The goal of this chapter was to take advantage of both the high speed of digital zoom and the ability of optical zoom to alter image detail and context.

When digitally zooming, the effect of image noise increases with digital magnification. It was shown that by adjusting the static measurement noise covariance of a Kalman Filter, the effect of noise could be attenuated considerably. However, those adjustments cause the performance of digital zoom to degrade in low noise situations. This unwanted side effect was removed by allowing the measurement noise covariance to change dynamically, so that it would scale with the measured image height. The resulting filter decreased the impact of image noise without noticeably degrading performance.

This chapter showed experimentally that adding digital zoom to an active vision system decreases fixation error while improving scale invariance.

In a hybrid-zooming system, camera focal length no longer affects image size. When also coupled with a dual-camera system, the focal length does not affect system stability. Therefore, focal length control algorithms can be developed to maximize image detail: a key to image recognition.

Chapter 5

Conclusion

5.1 Summary of this work

The goal of this work was to develop a system capable of capturing images of a specific size and position. To this end, a number of imaging concepts were introduced—including sensing, perception, the pinhole camera, projection models, and camera calibration—which provided a foundation for zoom understanding.

Preceding methods used a single-camera system to implement zoom which created a tradeoff between maintaining fixation and maximizing resolution—in the event of lost fixation, the system is unstable. To split the responsibilities, a second camera with fixed focal length was introduced. The symbiotic relationship between the cameras ensured that resolution goals could be met while maintaining overall system stability.

By placing the cameras side-by-side—giving them similar views—the images captured were comparable, differing only by focal length. Using digital zoom, further magnification synchronized the images, permitting interchangeability. Separately, the cameras' original images differ in resolution and context, but together the system benefits from both.

5.2 Future work

Two major contributions arose from this work: stability in the presence of zoom and the acquisition of scale invariant images. The possibilities of future work include the following:

5.2.1 Advanced tracking

The most logical extension of this work is to move beyond a simple tracker. While the color segmentation tracker worked as a proof-of-concept, the different characteristics of more advanced trackers could prove more pertinent, *i.e.* more relevant to the task at hand. For example, the CAMSHIFT algorithm has been shown to be successful in tracking faces in real-time [4]. The resulting images could be passed on to a higher level application, *e.g.* facial recognizer, eye gaze estimator, *etc.*

5.2.2 More cameras

With the dual camera zoom as described in Chapter 3, an overaggressive zoom results in a sudden decline in resolution. This decline can be less dramatic by using multiple cameras at various focal lengths.

5.2.3 Advanced zooming camera

Ideally, a system with no separation between the panoramic and zooming camera makes image splicing and camera correspondence trivial. Some promising developments have been shown in [32] which give optical zoom a free center of expansion and near-instantaneous magnification. The principles established in this work could be directly applied to that system.

Of course, the human eye is already equipped with a “zooming” camera: the foveal region. Eventually, technology will allow us to produce an active vision system equivalent to human sight—this will not be the final goal, but rather a milestone.

Appendix A

Miscellaneous

A.1 Derivation of constant velocity Kalman Filter

From [3]:

A constant-velocity target for a generic coordinate x is described by the equation

$$\ddot{x}(t) = 0 \quad (\text{A.1})$$

Since the position $x(t)$ evolves (in the absence of noise) according to a polynomial in time, such a model is also called polynomial and the resulting filter is called a *polynomial filter*.

In practice, the velocity undergoes at least slight changes. This can be modeled by the continuous time white noise $\tilde{v}$ as follows:

$$\ddot{x}(t) = \tilde{v}(t) \quad (\text{A.2})$$

where

$$E\tilde{v}(t) = 0 \quad (\text{A.3})$$

$$E\{\tilde{v}(t)\tilde{v}(\tau)\} = q(t)\delta(t - \tau) \quad (\text{A.4})$$

The state vector corresponding to A.2 is

$$\mathbf{x} = \begin{pmatrix} x \\ \dot{x} \end{pmatrix} \quad (\text{A.5})$$

In many application, the same model is used for each coordinate, and the motion along each coordinate is assumed “decoupled” from the others. The noises entering into various coordinates are also assumed mutually independent but with possibly different and time-varying intensities.

The continuous-time state equation is

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tilde{v}(t) \quad (\text{A.6})$$

where

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad (\text{A.7})$$

The discrete-time state equation with [uniform] sampling interval ΔT is

$$\mathbf{x}(k+1) = \mathbf{F}\mathbf{x}(k) + \mathbf{v}(k) \quad (\text{A.8})$$

where

$$\mathbf{F} = e^{\mathbf{A}\Delta T} = \begin{pmatrix} 1 & \Delta T \\ 0 & 1 \end{pmatrix} \quad (\text{A.9})$$

and the discrete-time process noise relates to the continuous-time version as follows:

$$\mathbf{v}(k) = \int_0^{\Delta T} T_0 e^{\mathbf{A}(\Delta T - \tau)} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tilde{v}(k\Delta T + \tau) d\tau \quad (\text{A.10})$$

The covariance of $\mathbf{v}(k)$, assuming q to be constant and using A.4, is

$$\begin{aligned} \mathbf{Q} &= \mathbf{E}[\mathbf{v}(k)\mathbf{v}'(k)] \\ &= \int_0^{\Delta T} T_0 \begin{pmatrix} \Delta T - \tau \\ 1 \end{pmatrix} \begin{pmatrix} \Delta T - \tau & 1 \end{pmatrix} q d\tau \\ &= \begin{pmatrix} \frac{\Delta T^3}{3} & \frac{\Delta T^2}{2} \\ \frac{\Delta T^2}{2} & \Delta T \end{pmatrix} \end{aligned} \quad (\text{A.11})$$

Bibliography

- [1] Brian D.O. Anderson and John B. Moore. *Optimal Filtering*. Dover, New York, 2005.
- [2] Shai Avidan. Support vector tracking. *IEEE Trans. on Pattern Recognition and Machine Intelligence*, 26(8):1064–1072, August 2004.
- [3] Y. Bar-Shalom and T.E. Fortmann. *Tracking and Data Association*. Academic Press, Inc., San Diego, CA, USA, 1988.
- [4] Gary R. Bradski. Computer vision face tracking for use in a perceptual user interface. *Intel Technology Journal*, (Q2):15, 1998.
- [5] Shao-Yi Chien, Shyh-Yih Ma, and Liang-Gee Chen. Efficient moving object segmentation algorithm using background registration technique. *IEEE Trans. Circuits Syst. Video Techn.*, 12(7):577–586, 2002.
- [6] Evan Clark. A multicamera system for gesture tracking with three dimensional hand pose estimation. Master’s thesis, Rochester Institute of Technology, 2006.
- [7] Robert Collins, Omead Amidi, and Takeo Kanade. An active camera system for acquiring multi-view video. In *Proceedings of the 2002 International Conference on Image Processing (ICIP ’02)*, pages 517–520, September 2002.
- [8] K. Daniilidis, Ch. Krauss, M. Hansen, and G. Sommer. Real-time tracking of moving objects with an active camera. *Journal of Real Time Imaging*, 1:3–20, 1998.
- [9] Daniel C. Dennett. *Freedom Evolves*. Viking Penguin, New York, NY, USA, 2003.
- [10] J. Denzler, M. Zobel, and N. Niemann. Information Theoretic Focal Length Selection for Real-Time Active 3-D Object Tracking. In *International Conference on Computer Vision*, pages 400–407, Nice, France, October 2003. IEEE Computer Society Press.

- [11] Jeffrey A. Fayman, Oded Sudarsky, Ehud Rivlin, and Michael Rudzsky. Zoom tracking and its applications. *Machine Vision and Applications*, 13(1):25–37, 2001.
- [12] David A. Forsyth and Jean Ponce. *Computer Vision: A Modern Approach*. Prentice Hall, New Jersey, 2003.
- [13] M. Greiffenhagen, V. Ramesh, D. Comaniciu, and H. Niemann. Statistical modeling and performance characterization of a real-time dual camera surveillance system. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 335–342, 2000.
- [14] C. Harris. Tracking with rigid models. In A. Blake and A. Yuille, editors, *Active Vision*. MIT Press, Cambridge, MA, 1992.
- [15] E. Hayman. *The Use of Zoom within Active Vision*. PhD thesis, University of Oxford, 2000.
- [16] E. Hayman, T. Thórhallsson, and D.W. Murray. Tracking while zooming using affine transfer and multifocal tensors. *International Journal of Computer Vision*, 51(1):37–62, January 2003.
- [17] Bernd Heisele, Purdy Ho, and Tomaso Poggio. Face recognition with support vector machines: Global versus component-based approach. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 688–694, 2001.
- [18] Intel Open Source Computer Vision Library. <http://www.intel.com/research/mrl/research/opencv/>.
- [19] Qian Huang, Yuntao Cui, and Supun Samarasekera. Content based active video data acquisition via automated cameramen. In *ICIP (2)*, pages 808–812, 1998.
- [20] R. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME, Journal of Basic Engineering*, 82:35–45, 1960.
- [21] Mengxiang Li and Jean-Marc Lavest. Some aspects of zoom lens camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(11):1105–1110, 1996.
- [22] Jonathan C. Rupe. Vision-based hand shape identification for sign language recognition. Master’s thesis, Rochester Institute of Technology, 2005.
- [23] Larry S Shapiro. *Affine Analysis of Image Sequences*. Cambridge University Press, New York, 1995.

- [24] Gideon P. Stein. Tracking from multiple view points: Self-calibration of space and time. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 1521–1527, 1999.
- [25] M. Sznaier, B. Murphy, and O. Camps. An lpv approach to synthesizing robust active vision systems, December 2000.
- [26] T. Camps O. Sznaier, M. Inanc. Robust controller design for active vision systems. *American Control Conference*, 3:2013–2017, 2000.
- [27] A. Murat Tekalp. *Digital video processing*. Prentice-Hall, Inc., 1995.
- [28] B. Tordoff and D. Murray. Resolution vs. tracking error: Zoom as a gain controller. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Madison, Wisconsin*. IEEE Computer Society Press, June 2003.
- [29] B. Tordoff and D. Murray. Reactive control of zoom while fixating using perspective and affine cameras. *IEEE Trans. on Pattern Recognition and Machine Intelligence*, 26(1):98–112, January 2004.
- [30] B. J. Tordoff. *Active Control of Zoom for Computer Vision*. PhD thesis, University of Oxford, 2002.
- [31] B. J. Tordoff and D. W. Murray. Reactive zoom control while tracking. Technical Report OUEL 2228/00, University of Oxford, 2000.
- [32] David V. Wick and Ty Martinez. Adaptive optical zoom. *Optical Engineering*, 43(1):8–9, 2004.
- [33] Reg Gordon Willson. *Modeling and calibration of automated zoom lenses*. PhD thesis, Carnegie Mellon University, Pittsburgh, PA, USA, 1994.
- [34] Ming-Hsuan Yang, David J. Kriegman, and Narendra Ahuja. Detecting faces in images: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(1):34–58, 2002.