Hindawi Publishing Corporation
Journal of Sensors
Volume 2016, Article ID 8137859, 13 pages
http://dx.doi.org/10.1155/2016/8137859

Research Article

Decoupled Multicamera Sensing for Flexible View Generation

Vivek K. Singh,[1,2] Danny Fernandes,[3] Mohan Kankanhalli,[4] and Thomas Haenselmann[5]

[1] *School of Communication and Infrormation, Rutgers University, New Brunswick, NJ 08901, USA*
[2] *MIT Media Lab, 75 Amherst Street, Cambridge, MA 02139, USA*
[3] *Cisco Systems India Pvt. Ltd., 10 M.G. Road, Bangalore 560 001, India*
[4] *School of Computing, National University of Singapore, Singapore 117417*
[5] *School of Applied Sciences, Hochschule Mittweida, Technikumplatz 17, 09648 Mittweida, Germany*

Correspondence should be addressed to Vivek K. Singh; v.singh@rutgers.edu

Received 27 August 2015; Accepted 17 November 2015

Academic Editor: Fanli Meng

Any sensing paradigm has three important components, namely, the *actor*, the *sensor*, and the *environment*. Traditionally, the sensors have been attached to either the actor or the environment. This restricts the kind of sensing that can be undertaken. We study a newer decoupled sensing paradigm, which separates the sensors from both the actor and the environment and tremendously increases the flexibility with which the scenes can be viewed. For example, instead of showing just one view, "how the environment sees the actor" or "how the actor sees the environment," a viewer can choose to see either one or both of these views and even choose to see the scene from any desired position in any desired direction. We describe a methodology using mobile autonomous sensors to undertake such decoupled sensing and study the feasible number as well as the placement of such sensors. Also, we describe how the sensors can coordinate their movements around a moving actor so as to continue capturing the required views with minimum overall cost. The practical results obtained demonstrate the viability of the proposed approach.

1. Introduction

There are three important components of any sensing paradigm, namely, the *actor*, the *sensor* (e.g., camera), and the *environment*. In the traditional video production paradigm, the sensors are part of the environment and the actor is observed. In robotics, on the other hand, the sensors are contained within the actor and the environment is observed. However, both these paradigms limit our observation to only one entity, either the actor or the environment. Recent trends such as selfie sticks, drones, and autonomous robotic cameras have paved the way for a novel sensing paradigm, which decouples the sensors from both the environment and the actor and allows observations of both of them.

This new paradigm of autonomous sensors capturing both the actor and the environment tremendously increases the flexibility with which the viewers can observe a scene. Detaching the sensors from the environment allows the viewers to see the environment itself. Furthermore, decoupling of the sensors from the actor allows the viewers to see the scene in any direction independently of where the actor is currently looking. Also, the *mobility* of sensors allows the images to be captured from multiple perspectives and ensures that moving actors can be handled seamlessly.

While this new paradigm can still handle the currently popular applications like surveillance and robot-sensing, it shall be especially useful in situations where currently prevalent paradigms fail. For example, interactive 3D television (3DTV) requires very high levels of interactivity between the viewer and the scene being observed. The users may want to have not only the "external vision," that is, observe how the world sees the actor, but also the "internal vision," that is, how the actor sees the world. For example, while watching the popular movie "Shrek," some viewers may watch it in the "default" mode of an external camera observing "Shrek." However, many viewers are also interested in seeing how "Shrek" sees the world through his "Ogre-Vision." In fact, a very popular show at Disneyland provides viewers a chance to experience this alternative view. However, this has been done only for a short movie which has been rendered artificially.

Also, the users may want to change the viewing angle dynamically on-the-fly and flexibly watch the show unconstrained by the director's preferences. Similarly, many sports enthusiasts would like to watch "Formula-1" car races from the racing tracks. While nowadays cameras are placed in the cars, the viewers are still able to watch the stadium/spectators only from one (or a few) specific angle(s) as dictated by the car's physical position. With our novel sensing paradigm, the viewers can freely choose to see what one can see "if he were at position X looking in direction Y."

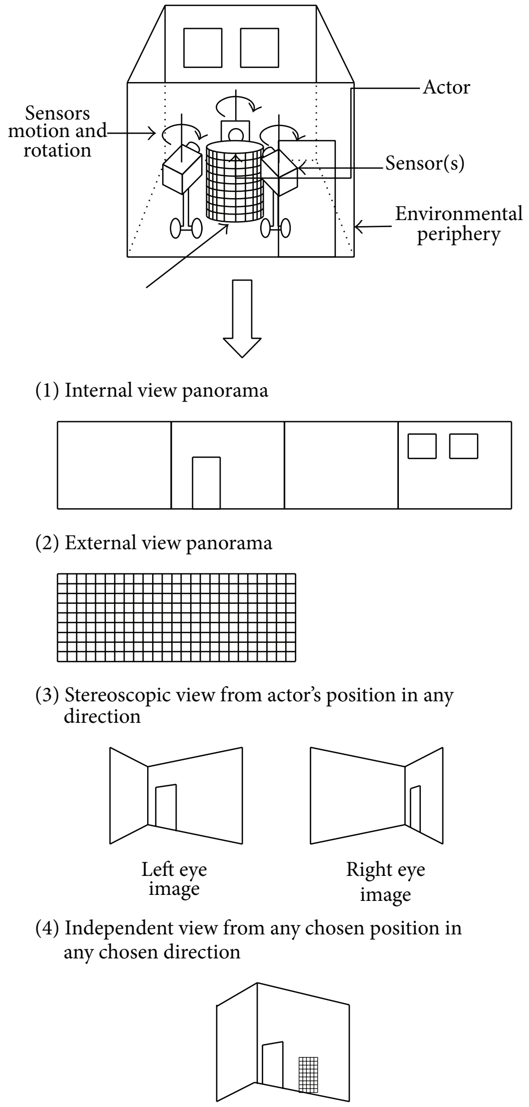
Thus we propose a novel sensing strategy to flexibly generate views consisting of both the actor and the environment. As shown in Figure 1, we want to allow the viewer to choose from four different types of views, namely, internal view mosaic (i.e., "environmental panorama view as seen from the mid actor"), external view mosaic (i.e., "an unwrapped image of the actor itself"), a stereoscopic view from actor's perspective looking in any direction (independent of where the actor is actually looking), and lastly an independent view (i.e., looking from any position X in direction Y). However, in order to generate these views, we need to handle two specific challenges. Firstly, we need to find the appropriate positions where the cameras can be placed around (static or moving) actor such that 360° images of the actor and the environment can be captured. Next, we must use methods like view-morphing and stitching which can combine these captured images in order to generate the required views.

In this paper, our aim is to firmly ground the decoupled sensing paradigm and then focus on solving the first problem, that is, finding the optimal placement and coordination strategy for mobile cameras moving around an actor to continuously capture 360° images of the actor and the environment. The related problems of image stitching and dynamic view generation based on user choice from multiple cameras have been discussed by an array of recent academic works, for example, [1–3], and are increasingly becoming accessible in consumer facing technology, for example, [4, 5].

While the decoupled sensing paradigm is generic and can include both ground-based and aerial sensors, here we focus on the grounded sensors to validate the proposed concepts. Hence, for capturing the required images, we use a group of custom-made autonomous cameras (as shown in Figure 2), each of which can place itself at an appropriate position with respect to the main actor and undertake a 360° (or their maximum pan angle) rotation along their own axis and then move around the actor in a specific pattern. This procedure allows us to obtain a compilation of images to be used for dynamic scene creation at a later time. Specifically, we analyze the feasible regions, where these cameras can position themselves, their coordination strategy for capturing the images of a moving actor and the resulting trade-offs in the number of cameras, the image quality, and the sensing delay incurred. We also find the minimum number of sensors required for undertaking such a sensing task.

To demonstrate the feasibility of the proposed approach, we provide mathematical analysis, simulation results, and practical view generation examples.

The organization of the remainder of this paper is as follows. We describe the related work in Section 2. The proposed method for finding camera positions and their coordination is described in Section 3. Section 4 describes the various evaluation results and Section 5 gives the conclusions.



(1) Internal view panorama

(2) External view panorama

(3) Stereoscopic view from actor's position in any direction

(4) Independent view from any chosen position in any chosen direction

Figure 1: Types of flexible views available to the viewer.

2. Related Work

Kelly et al. in their multiperspective imaging work [6] have described an interactive method for users to view a real-world environment. The approach of letting a user choose his own view and look in any direction is very interesting and has been incorporated in our current work. Kanade [7] has also described the use of multiple cameras placed at stadium periphery to create 360° views with respect to any chosen player. However, both these works focus on creating a photorealistic 3D model of the scene and then rendering virtual views, while we make use of mobile sensors and decoupled sensing to capture the views from multiple perspectives. The mobility of sensors in our approach allows

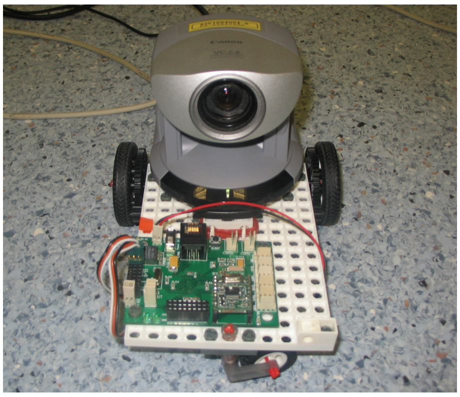


Figure 2: One mobile sensor.

wider variety of perspective images to be captured (rather than being rendered) and reduces the computational cost in generating the required user views.

Borovikov et al. have described a multiperspective imaging framework in [8]. They employ 64 cameras for allowing multiple views of the presented scene. Similarly, Seitz and Kim [9] describe stereoscopic view generation from multiple perspectives. However, both these works have cameras placed at the environment periphery and do not consider the actor-centric view of the environment which can be undertaken by our detached sensing paradigm. Recent products like Google Jump [4] and Facebook 360 [10] are creating consumer facing tools to allow for viewing from different positions. However, the considered cameras are fixed in circular and are not able to move around the room independently as considered in this work.

Zhang and Chen [11] have described an "Active Rearranged Capturing" approach to progressively improve the quality of rendered views as obtainable using image based rendering. They use a grid of 48 cameras, each of which has (limited) capability for readjustment to improve the quality of image capture. The ideas for image based rendering and progressive realignment are interesting and we have proposed similar strategies in our work. However, we also handle "actor-centric" views via our decoupled sensing paradigm. Furthermore, we allow the mobile sensors to move freely in the space and have 5 degrees of freedom (Pan, Tilt, Zoom, and movement(s) in x-axis and y-axis) as opposed to 2 provided (panning and side-stepping movement in x-axis) in their system.

In multirobotic imaging, there has been some interesting work on obtaining object-of-interest images using multiple robots. For example, Parker [12] has studied how moving objects can be tracked using cooperative robots. Similarly Gerkey and Matarić [13] have studied an auction based mechanism for coordinating robots in undertaking tracking tasks. However, these (and similar robotics) works focus on obtaining just one image for the object of interest from any angle. On the other hand, our problem needs optimal image capture from multiple perspectives which can be employed for ex-post 2.5D image production as well as internal and external panoramic view creation.

Yang et al. [14] describe a method for fast panoramic face mosaicing. In particular, they describe a method of creating face mosaics using 5 static cameras in different angles. Their work is interesting inasmuch as it advocates exploitation of newer techniques for panoramic facial image creation. However, their work focuses only on (an equivalent of) external image mosaicing and does not exploit the use of moving cameras for better positioning across time instances.

Thus, on the whole, while we find significant attention has been given to multiperspective imaging by the research community, it has mostly been undertaken via (computationally heavy) 3D modeling using sensors placed at the environment periphery. We, on the other hand, propose to undertake similar tasks using a sampling based method which uses decoupled mobile sensors followed by comparatively cheaper processes like stitching and morphing.

Thus, the key contribution of this paper is to introduce this decoupled sensing paradigm which tremendously increases the flexibility in view generation while still keeping the computational costs low. We also study the optimal placement and selection of sensors for undertaking sensing tasks using this paradigm which are indeed being studied for the first time due to the novel problem definition.

3. Multiperspective Imaging Using Autonomous Mobile Cameras

The process of capturing the requisite images consists of two separate parts. Firstly, the cameras must organize themselves around the actor to form a sensing structure which appropriately capture the entire circumference of the actor and the environment. In our approach we obtain "feasibility regions" for each camera which ensure that the cameras, when placed within their respective "feasibility regions," can indeed combine to provide the required images.

The second part considers how the cameras can reorganize themselves around a moving actor such that they continue to capture the required views. It also includes how images can be captured from multiple perspectives to reduce distortion in generated views while still minimizing the camera movement cost. In our approach, we observe the actor's movement and also consider how we can reconfigure the camera positions to improve the quality of overall generated views. This results in new sets of feasibility regions for each camera. Next, we calculate the costs required for each camera to move to any of the newly established feasibility regions. We then try to optimally allocate the cameras to these feasibility regions so that the overall reconfiguration cost is minimized.

For ease of understanding let us decouple the two (sub)problems as of now, and we first focus on how cameras can form an optimal sensing structure around a *static* actor (or any other object of interest) in order to fulfill requisites for the various required views.

3.1. Static Actor Case. Let us consider the case where we have one actor who is surrounded by n cameras trying to capture the required images. As mentioned before, the basic requirements for all view creations are 360° images of the actor

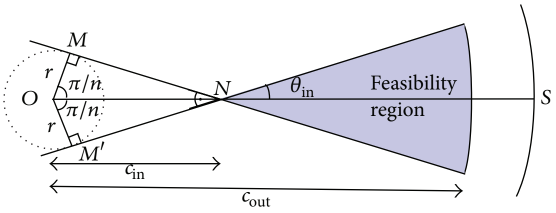


Figure 3: Requirements for actor image capture.

and the environment. Besides, to support stereoscopic view of any part of the environment, any point on the environmental periphery must be captured by two distinct cameras. Thus, each of these requirements poses certain restrictions on the positioning of the cameras, thus resulting in "feasible" and "unfeasible" regions. We proceed to find these "feasibility regions" based on each of the individual requisites and then later combine them to obtain overall "feasibility regions" (or "FR"s). Please note that, for ease of formulation, we assume the actor and the environment to be cylindrical and convex. Furthermore, we concentrate on the single actor case and neglect any occlusions caused by the cameras themselves onto the other cameras.

3.1.1. Requisite 1: Capturing the Entire Actor Circumference. This requirement enforces that each of n cameras must cover at least $2\pi/n$ angle on the actor's external circumference as shown in Figure 3 so that the entire circumference is collectively covered. In the figure, O represents the center point of a circle of radius r which represents (or circumscribes) the actor. If (without the loss of generality) we choose a specific direction as the base axis, we can split the problem setup into two symmetric halves. In each half (let us call them "top" and "bottom"), the camera must capture images of at least π/n angle as also shown in Figure 3. This results in a minimum distance which must be maintained between the camera and the actor. This limiting condition occurs where a pair of tangents making an angle of $2\pi/n$ meet. This is because any point nearer shall form a smaller angle at the actor circumference and thus a camera placed at that point cannot capture the required images irrespective of the pan angle or field of view available to it.

Thus, by considering ΔMNO, we can find the minimum distance to be maintained (c_{in}) as

$$c_{\text{in}} = r \times \sec\left(\frac{\pi}{n}\right). \tag{1}$$

The tangents MN and $M'N$ also form a restriction on the maximum angular displacement in the camera position. In effect,

$$\left\|\theta_{\text{in}}\right\| < \frac{\pi - 2\pi/n}{2}, \tag{2}$$

where θ_{in} is the angular displacement of the physical camera position with respect to the "ideal" position. This is due to the fact that a camera positioned below the line MN shall not be able to capture images of point M which is required for the appropriate completion of requisite 1. Similarly, any point above $M'N$ shall not be able to capture point M'.

Finally, there is also a bound on the maximum distance possible between the camera and the actor as the actor images need to be captured at a minimum required resolution. This distance depends on the zoom capability of the camera and we call it c_{out}. Thus the FR for requisite 1 shall be a sector bounded by parameters c_{in}, θ_{in}, and c_{out}.

3.1.2. Requisite 2: Capturing the Entire Environmental Circumference. In order to cover the entire outer circumference, each camera must cover at least one $2\pi/n$ angular arc (or two π/n angular arcs) on it. In our proposed approach, the cameras must capture the images of the actor and the environment in the same sensing cycle; hence, it makes sense for cameras to capture both in one panning action as shown in Figure 4(a). This results in two symmetric halves where the camera captures the environmental images. Further details for the "top" half have been shown in Figure 4(b), where a camera placed at point P captures images of the actor centered at O and then continues its pan motion in order to obtain images for the arc LQ on the environmental periphery which forms an angle π/n on O. We assume actor radius (r), distance between actor and the environment (E), and the (symmetric half of) camera "net pan angle" (ϕ) to be known. Also, ϕ is taken to be the combination of the camera pan angle (ϕ_{pan}) and camera field of view (ϕ_{fov}).

Clearly, the further the camera goes from the center of the actor the more angle it can capture. But there exists a limit to how near it can come to the actor. Coming nearer to the actor might be useful for reducing the movement cost for rotation vector between capture cycles (as further discussed in next section) and increasing the resolution of the actor images being captured.

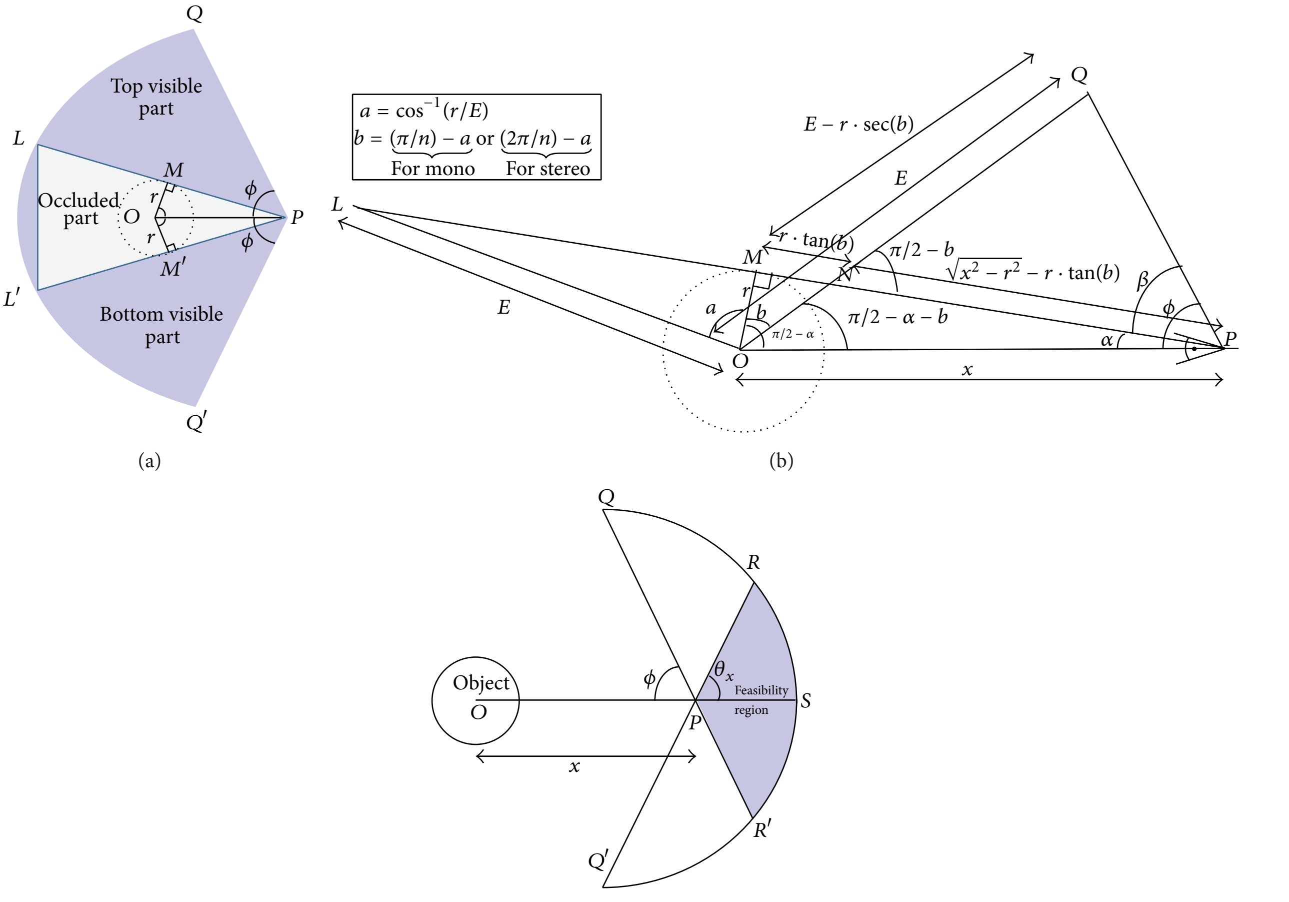
Please note that as the camera moves further away from the actor the occluded angle (i.e., α as shown in Figure 4) keeps decreasing and the angle for which the environment is captured (i.e., β) keeps increasing. There is a point P beyond which β can capture larger than the required angle (π/n) at O. We can find the distance x for this point P by first formulating the values of α and β in terms of x and other known environmental variables and then solving the converse problem.

By considering ΔMPO, we can find the value of α as

$$\alpha = \cos^{-1}\frac{r}{x}. \tag{3}$$

By solving ΔMNO and then formulating equations in ΔNQP, we find the value of β as

$$\beta = \tan^{-1}\left\{\frac{(E - r\cdot\sec(b))\cdot\cos(b)}{\sqrt{x^2 - r^2} - r\cdot\tan(b) - (E - r\cdot\sec(b))\cdot\sin(b)}\right\}. \tag{4}$$



(a)

(b)

(c)

Figure 4: Environmental image capture.

We know that ϕ is the sum of the occluded (α) and the unoccluded (β) angles. Thus, to solve the converse problem, that is, distance x in terms of α and β, we formulate

$$\phi = \cos^{-1}\frac{r}{x} + \tan^{-1}\frac{k_1}{\sqrt{x^2 - r^2} - k_2}, \qquad (5)$$

where $k_1 = (E - r \cdot \sec(b)) \cdot \cos(b)$ and $k_2 = -r \cdot \tan(b) - (E - r \cdot \sec(b)) \cdot \sin(b)$.

Note that k_1 and k_2 are known values independent of x and thus the only unknown in (5) is x.

This equation can be converted into the standard quadratic form if we use $\tan(\cdot)$ operator on both sides and define a new unknown variable u as $u = \sqrt{x^2 - r^2}$. Thus,

$$u = \frac{-\cos\phi}{2 \cdot (r \cdot \sin\phi - k_2 \cdot \cos\phi + k_1 \cdot \sin\phi)} + \frac{\sqrt{\cos^2\phi - 4 \cdot (r \cdot \sin\phi - k_2 \cdot \cos\phi + k_1 \cdot \sin\phi) \cdot (k_2 \cdot r \cdot \sin\phi - k_1 \cdot r \cdot \cos\phi)}}{2 \cdot (r \cdot \sin\phi - k_2 \cdot \cos\phi + k_1 \cdot \sin\phi)} \qquad (6)$$

and finally x can be found as

$$x = \sqrt{u^2 + r^2}. \qquad (7)$$

We can cover the required angle at the outer circumference if we place the camera at any point on the base axis at a distance greater than x. The amount of lateral movement the camera is allowed is dictated by the angle formed between the base axis and the last point on the outer circumference that needs to be covered, that is, point Q. Thus, in effect, this value depends on the symmetric half of camera net pan angle (ϕ). As shown in Figure 4(c), if the camera moves in front of line QR', it cannot capture the environmental circumference in the top half till the required angle. Similarly, if it moves in front of line $Q'R$, it cannot capture the lower circumference till the required angle.

This results in sector-shaped feasibility region $PRSR'$ formed at angles $\pm\theta_x$ starting from point P at distance x on the base axis and ending at a point on the outer circumference S. An interesting point to note is that the angle θ_x for this

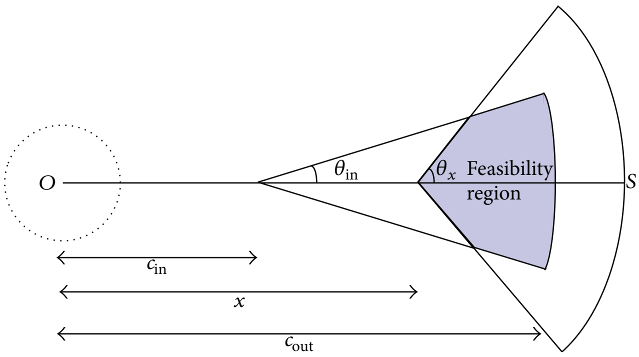


Figure 5: Overall feasibility region.

sectoral shape shall always be acute irrespective of whether ϕ is acute or obtuse. This is so because acute ϕ will cause acute θ_x limit in opposite half (e.g., $\angle QPO$ will cause a corresponding constraint on $\angle SPR'$), while obtuse ϕ will cause acute θ_x restriction in its own half.

Thus, the limits on the value of θ_x can be defined as

$$\|\theta_x\| < \min\{\phi, \pi - \phi\}, \tag{8}$$

where π is indeed the theoretical limit of ϕ as we are only considering one half of the symmetric setup.

3.1.3. Requisite 3: Stereoscopic Imagery. For obtaining stereoscopic images of the outer circumference, each point on it must be covered by two cameras and hence each camera must cover twice the angles as compared to simple monoperspective imaging. The way we have formulated the problem in previous requisite is generic and if we just replace the angle to be covered at the center to be $4\pi/n$, that is, $2\pi/n$ in both up and down directions, we can use the same set of derivations as above to obtain the FR. Hence, the only parameter that changes in Figure 4(b) is angle b whose value now shall be $(2\pi/n) - a$.

The restrictions on feasible region posed by each of these requirements can also be combined to obtain an overall FR. An illustration of such combination is shown in Figure 5. Please note that since stereoscopic imagery requirement "subsumes" the environmental capture requirement, we consider only the more restrictive requisite to calculate the overall FR. The overall FR shall be at the intersection of the two sector-shaped feasibility regions and is parameterized by c_{in}, c_{out}, θ_{in}, x, and θ_x. The angular position of this FR shall indeed be different for each camera, as each of these cameras needs to be placed at equal angular distances around the actor in order to achieve the overall task(s). Also, without the loss of generality, if we fix one camera of the cameras as the "base" camera, we can assign its angular direction as zero. The remaining camera's direction can be calculated with respect to this base camera.

3.1.4. Minimum Requirements Analysis. Based on the various requirements discussed in this section, we can find the minimum number of cameras and their required net pan angle for successful capture of required images. Let us first try to find the minimum number of cameras based on θ_{in} requirement. As mentioned in (2), θ_{in} should be less than $(\pi - 2\pi/n)/2$ and it must have a positive value. This equation translates to

$$n > 2 \tag{9}$$

which means the minimum number of cameras required is 3. This number, 3, also is able to meet the constraints of c_{in} and so forth and thus qualifies as the minimum number for the tasks being considered. The minimum net pan angle required for the cameras is dictated by the environmental capture requisite. The bounding condition happens when $E \gg r$ and α can be neglected. In such a case, for monoperspective viewing, the minimum (symmetric half of) net pan angle required is

$$\phi > \frac{\pi}{n} \tag{10}$$

and for stereoscopic viewing it is

$$\phi > \frac{2\pi}{n}. \tag{11}$$

Thus, for 3 cameras, the minimum net pan angle required for stereoscopic viewing is 240°.

3.2. Dynamic Camera Reconfiguration for Moving Actor. As the actor moves, the cameras must also dynamically reorganize themselves such that they continue to capture the actor images at same (or better) quality to support various view generation. An overview of our proposed approach to undertake this reorganization from sensing cycle k to $k+1$ is as follows. We first translate the FRs found in cycle k based on actor's displacement to find potential FRs in cycle $k+1$. Next we rotate these newly found FRs based on a "progressive realignment" strategy to (potentially) improve view quality. This process of feasibility region translation and realignment has been shown in Figure 6. Lastly, we assign these computed FRs to the individual sensors in an optimal manner. Such a process is repeated for each sensing cycle. Please note that we assume that actor and camera positions are known or easily obtainable in each cycle. Now, let us look at each of these steps in more detail.

3.2.1. Step 1: Simple Displacement Based on Actor Movement. The first step is the simple translation of FRs to follow an actor who is moving across the "stage." The translation vector can simply be defined as

$$D_{\text{tr}} = [x_{k+1} - x_k, y_{k+1} - y_k]^T, \tag{12}$$

where $[x_{k+1}, y_{k+1}]$ and $[x_k, y_k]$ are the actor positions in cycles $k+1$ and k, respectively.

3.2.2. Step 2: Progressive Realignment for Reducing Image Distortion. The simple displacement described in step 1 shall result in zero (or almost zero) relative movement between

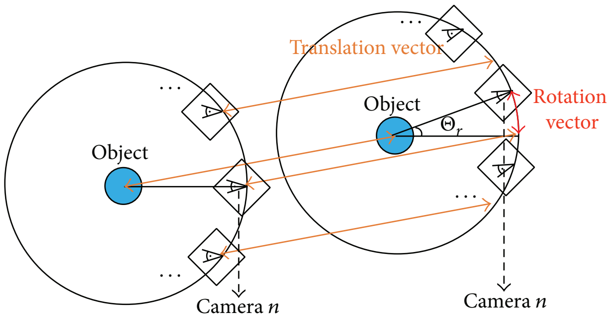


FIGURE 6: Consecutive configurations for the sensors.

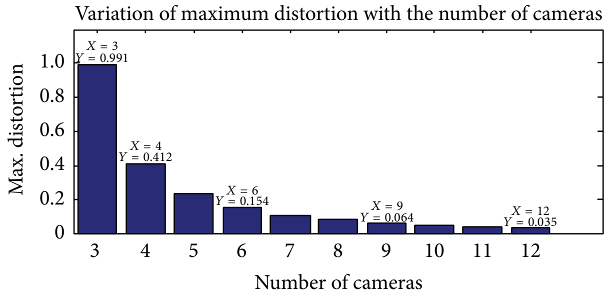


FIGURE 8: Distortion versus number of cameras.

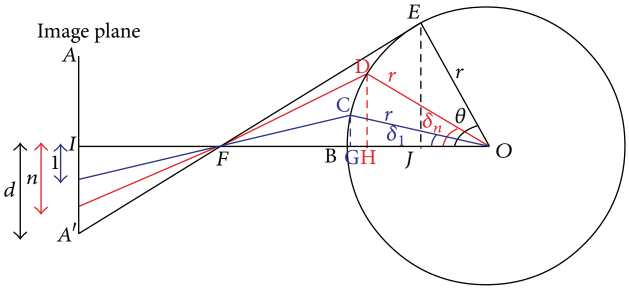


FIGURE 7: Distortion analysis.

the camera structure and the actor. This means that there will also be zero *improvement* in the overall quality of the generated views. Thus, we employ a "progressive realignment" strategy to improve the view quality by increasing the number of perspectives from which the image(s) are being captured. The basic idea is as follows. If 4 cameras were obtaining actor images with an angular distance of 90° between them, we could move the cameras by 45° each so that they can create newer perspectives for image capturing. Thus, if the images from these two capture cycles were combined, we could use images from 8 different perspectives differing from each other by 45°, which shall reduce the *distortion* in the final views generated.

To quantify this notion of distortion, let us consider (top view) Figure 7 where the actor with center O is being captured by a camera with image plane AA' and focal point F. Let the camera's horizontal resolution be $2d$ pixels. Each part of the arc BCE is captured by some pixels on the camera plane AA'. However, the relative lengths of the arcs being captured vary depending on the angular position of the arc. While a relatively small distance of arc is covered by one pixel near the axis, a much larger distance is covered further away. For the purpose of our analysis we define distortion as the ratio of the arc lengths captured by one pixel near the axis to that at a disparate angular position. In Figure 7, using similar triangles, we can formulate the following equation:

$$\frac{r\sin\delta_1}{r\sin\theta} = \frac{1}{d}. \tag{13}$$

We can obtain similar equations for any chosen pixel, say n:

$$\frac{r\sin\delta_n}{r\sin\theta} = \frac{n}{d}. \tag{14}$$

To calculate the distortion at any particular angular position, we compare the arc length captured by the first pixel with that captured by the nth pixel and quantify the difference as distortion. As the arc lengths created are directly proportional to the angle they subtend at the center O, we in effect need to calculate the ratio between the angles which correspond to the 1st pixel and the nth pixel. Note that, in the above formulation, the incremental angle corresponding to nth pixel is represented by $\delta_n - \delta_{n-1}$.

Thus, the distortion at any point is

$$\begin{aligned}\text{dist}_n &= \frac{\sin^{-1}\left(\left((n+1)/d\right)\times\sin\theta\right) - \sin^{-1}\left((n/d)\times\sin\theta\right)}{\sin^{-1}\left((1/d)\times\sin\theta\right)} \\ &\quad - 1\end{aligned} \tag{15}$$

and the maximum distortion is

$$\text{dist}_{\max} = \frac{\theta - \sin^{-1}\left(\left((d-1)/d\right)\times\sin\theta\right)}{\sin^{-1}\left((1/d)\times\sin\theta\right)} - 1. \tag{16}$$

Using this function we plotted the value of maximum distortion (as shown in Figure 8) for different number of cameras. We found that $\text{dist}_{\max}$ is 99% for $n = 3$ and comes down exponentially to approximately 15% and 5% with $n = 6$ and $n = 9$, respectively. Thus, if we can employ just two (resp., 3) rounds of realignment for one image combination phase, the distortion ratio can be brought down drastically.

The precise value for realignment of FRs during such combination phases is based on a method inspired by incremental JPEGs. We capture the images from n perspective positions in the first cycle. This provides a base case quality for creating the required views. We then rotate the cameras such that they maximally try to increase the number of perspectives in a binary sort equivalent manner (e.g., $0.5 \times 2\pi/n$ for round 2 and so on). Thus we can get "rough" equivalents of $n \times k$ cameras after k rounds. The initial few rounds carry more coarse level information and each iterated level can be used to fine-tune it. This coarse-fine approach assumes that the distance traveled by actor between

$$\begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,n-1} & c_{1,n} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,n-1} & c_{2,n} \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ c_{n,1} & c_{2,n} & \cdots & c_{n,n-1} & c_{n,n} \end{bmatrix}$$

(a)

$$\begin{bmatrix} c'_{1,1} & 0 & \cdots & 0' & c'_{1,n} \\ c_{2,1} & 0' & \cdots & c'_{2,n-1} & c'_{2,n} \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0' & c'_{2,n} & \cdots & c'_{n,n-1} & c'_{n,n} \end{bmatrix}$$

(b)

$$\begin{bmatrix} c'_{1,1} & 0 & \cdots & c'_{1,n-1} & c'_{1,n} \\ c_{2,1} & 0' & \cdots & c'_{2,n-1} & c'_{2,n} \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0' & c'_{2,n} & \cdots & c'_{n,n-1} & c'_{n,n} \end{bmatrix}$$

(c)

$$\begin{bmatrix} \cancel{c'_{1,1}} & \cancel{0} & \cancel{\cdots} & \cancel{c'_{1,n-1}} & \cancel{c'_{1,n}} \\ \cancel{c_{2,1}} & \cancel{0'} & \cancel{\cdots} & \cancel{c'_{2,n-1}} & \cancel{c'_{2,n}} \\ \vdots & \cancel{\ddots} & \ddots & \vdots & \vdots \\ 0' & \cancel{c'_{2,n}} & \cdots & c'_{n,n-1} & c'_{n,n} \end{bmatrix}$$

(d)

FIGURE 9: Kuhn's Hungarian method for OAP.

cycles is very small as compared to the overall area. It also assumes that the environment periphery does not change significantly between sensing cycles and the actor undergoes only translation motion and that shape deformation and so forth can be neglected. We use this incremental coarse-fine approach for one combination phase during the period for which each of these conditions can be satisfactorily met. If any of changes exceed a threshold value, we start a new combination phase.

Note that this analysis is also useful for system designers to make informed decisions about trade-offs between adding more cameras to reduce capturing time and allowing more delay (i.e., more realignment steps) to maintain low cost for any acceptable level of distortion.

3.2.3. Step 3: Optimal Assignment of Cameras to FRs. The final step is to optimally assign the cameras to these newly found FRs such that the overall movement cost is *minimized* where the cost refers to the sum of Euclidean distances that need to be traveled by the cameras and the inertial cost (cost of starting movement as opposed to remaining stationary) where applicable. To undertake such an assignment, we first need to calculate the costs for each camera to move from its current position (say in iteration k) to *any* of the n feasible regions in iteration $k + 1$. We do not enforce that camera i should only move to feasible region i in cycle $k + 1$, as "dynamic role swapping" could be useful to reduce the overall cost. For example, consider a two-camera case where each camera incurs some inertial cost (the fact that it must move) followed by some movement cost in this reconfiguration. It may so happen that camera 1 falls into FR2 as computed in iteration $k + 1$ (i.e., FR_2^{k+1}). Hence, it would make sense for camera 1 to handle FR_2^{k+1} and save the inertial cost, even though camera 2 needs to travel a bit more.

To formulate this problem generically, we have n cameras, each of which needs to move to one of the n FRs and there is a cost involved for any camera (say i) to move to any of the the FRs (say j) which can be represented as $c_{i,j}$. Thus, if we represent the set of all cameras as C and that of all feasibility regions as F, our aim is to find the allocation set a^* out of all allocation sets $C \times F \rightarrow A$ which minimizes the overall cost to be incurred. Thus, we need to find

$$\min \sum_{a \in A} c\{f(a)\}, \tag{17}$$

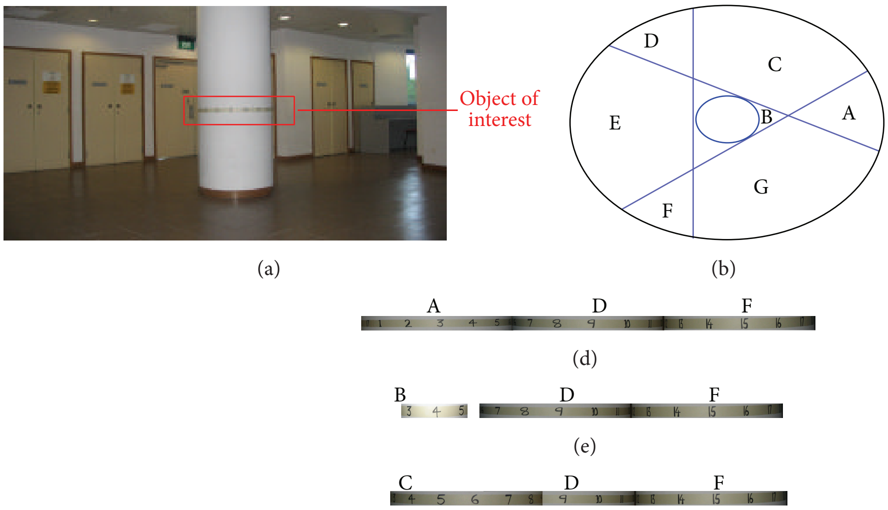
where $c\{f(a)\}$ is the sum of costs for various individual camera-task assignments for allocation a.

This problem can be translated to the "Optimal Assignment Problem" (OAP), wherein n agents need to be assigned n tasks in an optimal manner. If we represent each camera's costs for moving to each of the FRs as

$$C_{i,j} = \begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,n} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,n} \\ \cdots & \cdots & \cdots & \cdots \\ c_{n,1} & c_{n,2} & \cdots & c_{n,n} \end{bmatrix}, \tag{18}$$

we can find the optimal assignment solution using Kuhn's Hungarian method [15].

The key idea of Kuhn's method is as follows. We represent the costs as shown in Figure 9(a). Next, we find the least cost for each task (i.e., minimum value in each row) and subtract the entire row by it. This results in at least one nonzero element per row as (partially) shown in Figure 9(b). If multiple zeroes exist, the allocation must try to find a combination which results in no conflicts between rows as each agent can be assigned only one task. If no such combination is possible (as in Figure 9(c) where rows 1 and 2 are conflicting in column 2), the idea is to iteratively mark out all the rows and columns involved (as in Figure 9(d)) and then find the 2nd least-costing alternative for one of the conflicting rows as this shall provide the best overall *feasible* result. In fact, the method is proven to give guaranteed optimal solutions in $(O(n^3))$ polynomial time [15], and hence we employ it to find the optimal camera allocation for dynamic reconfiguration at each step.



Combination	Average score (max.: 10)
A, D, F	8.8
C, D, F	7.2
G, D, F	7.0
B, D, F	5.4
E, D, F	3.2

(a) (b) (c)

(d)

(e)

(f)

Figure 10: Feasibility region analysis.

4. Evaluation

4.1. Creation of Actor and Environment Panorama. Our first experiment aims to verify the premise that there exist specific feasibility regions images captured from which satisfy the necessary conditions for view generation (and conversely those from outside these FRs do not). To authenticate this hypothesis we chose a cylindrical pillar as an "actor" or the "object of interest" and studied the effect of capturing images from various points around it, given that our aim is to capture the entire circumference of the pillar. As discussed earlier, capturing the entire circumference is a basic requirement to facilitate the views of any point of the "actor" as may be desired by the viewer ex-post.

To make the results clearer we pasted an adhesive tape marked with numbers 1 through 17 at 15 cm distance each as shown in Figure 10(a) on the pillar and then took the images from various points around it as shown in a (top view) representation in Figure 10(b). Each of the alphabets A through G represents an image capture position and it is clear that some of the capturing points fall into the 3 feasibility regions (namely, A, D, and F) while others do not. To compare the results, we chose points D and F as reference points and permuted the third point from amongst the remaining points. The images from the chosen points were simply padded as shown in Figures 10(d), 10(e), and 10(f). We can clearly see that while Figure 10(d) shows the entire circumference with all numbers 1 to 17 visible, any other combination fails to do so. For example, Figure 10(e) shows the combination of images captured from points B, D, and F. As B is too close to the pillar (and hence outside the FR), it cannot capture the required part of circumference. Similarly, image captured from point C in Figure 10(f) is outside the FR and cannot capture the required view, thus resulting in incomplete panorama.

The subjective evaluation results by 5 viewers from our lab regarding the reconstruction completeness (neglecting distortion) also corroborated the obtained results. As shown in Figure 10(c), the images obtained from the only combination which covered the required feasibility regions (i.e., A, D, and F) obtained the highest average points (i.e., 8.8/10). The others which did not cover the feasibility regions obtained significantly lesser results. These results helped us verify that feasibility regions do exist and the images captured from them can provide a much better 360° reconstruction than those from outside them.

After conceptually verifying the notion of FRs, we next proceeded to capture images using three mobile sensors around an actor. Our aim was to capture images to facilitate both internal and external panorama of the scene. The panoramas can be used as an evidence that the entire circumference has been captured and that if need be the viewer can be presented with parts of the panorama with view-morphing so as to simulate him looking from any particular direction onto any particular angle.

One sample setup in our lab with three mobile sensors has been shown in Figure 11. The mobile sensors are actually Canon VC-C4 cameras placed on custom-built mobile vehicles which are controlled by a PC using bluetooth connection. We used the cameras to capture the images of a "pokemon" actor and the lab environment. The images were captured using the concept of feasibility regions and the obtained results for actor as well as the environment panorama have been shown in Figures 12 and 13, respectively.

While the details for the fourth view generation (i.e., independent view from an arbitrary position and angle) are left outside the scope of this paper, we present its preview at a supporting website [16], a snapshot of which is shown in Figure 14.

4.2. Comparison between Optimal Assignment and Baseline Method. To study the effects of FR and optimal assignment on the sensor movement cost we calculated the trajectories

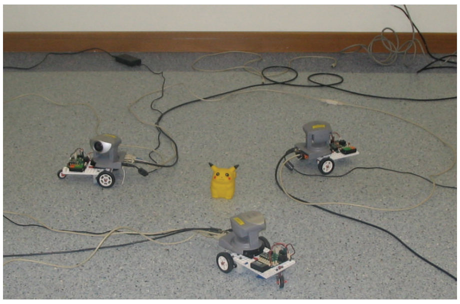


Figure 11: One sample setup.

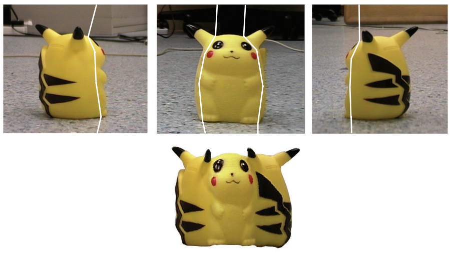


Figure 12: Actor mosaic.

and costs for a coordinated sensing scenario using Matlab. We computed the trajectories (and costs) firstly for the baseline approach and then compared it with a feasibility region (FR) based approach and finally with the optimal assignment cum feasibility region (OAP-FR) based approach. The cost in our study refers to the sum of Euclidean distances traveled by individual cameras and the inertial cost, that is, the transition cost for cameras to start moving. This transition cost was set as equal to 2 units of distance in our study.

To be fair to the baseline approach, we assumed that the cameras are continuously placed at c_{in} (as described in Section 3.1) position which incurs least translation cost while still being feasible. However, they moved only from one such calculated point at time k (c_{in}^{k}) to the next calculated single feasible point c_{in}^{k+1}. FR based approach on the other hand allowed the camera (say i) to move to any position in the calculated FR for time $k+1$ (FR_{i}^{k+1}). This meant that it could choose to move to the nearest point of the FR and does not need to travel to the specific point c_{in}^{k+1}. Furthermore, in certain situations, it may not need to move at all, as it could automatically fall into its FR for the next cycle. Lastly, in the OAP-FR approach, we allowed the cameras to calculate the costs to move to *any* of the FRs and an optimal assignment of cameras to FRs was chosen which minimizes the overall cost. Thus, camera i could fall into $i+1$th FR (or any other FR for that matter) and not move at all if doing so minimized the overall cost.

The calculated trajectories (only the first 4 cycles have been shown for clarity) for the different approaches have been shown in Figure 15. Figure 15(a) shows the movement of the actor (with actor center marked as ◯), starting at $[0,0]$ and moving $[2,2]$ in each cycle. Figure 15(b) shows the trajectories taken by the baseline method wherein the minimum distance c_{in} is 10. The cameras follow the translation as dictated by the actor movement and then undertake a 60° rotation to increase the number of sensing perspectives. We fix the maximum allowed sensing delay to be 2 cycles. The starting position of each camera has been marked with its respective number and its movement has been shown in the figure. Similarly, Figures 15(c) and 15(d) show the camera movement based on FR and OAP-FR based approaches.

The costs for the three approaches over a period of 100 sensing cycles have been shown in Figure 16. As can be seen, the cost for the FR based method is lesser (3369 units after 100 cycles) than the baseline method (3624 units after 100 cycles). This is due to the fact that the FRs allow a larger flexibility for the camera to move and it is not constrained to move to a specific point. Similarly, the cost for OAP-FR based approach is significantly lesser (1807 units after the 100 cycles) than the baseline approach as it allows the cameras to move to any specific FR. This "dynamic role swapping" between cameras also increases the probability of cameras falling into some FR and saving the inertial movement cost. For example, as shown in Figure 15(d), camera 3 needs to actually move only once in 4 cycles as it can take over the FRs for camera 2 and camera 1 for one cycle each and thus reduce the overall cost.

4.3. Effect of Multiple Perspectives on Quality of Images. Our next aim was to study the effect of multiple perspectives on the quality (or conversely distortion) of the images captured. We compared the distortion in the 360° image of the "actor" discussed in Section 4.1 for images reconstructed using 3, 6, and 12 with a ground truth "flat" image. Figures 17(a), 17(b), and 17(c), respectively, display samples of the image as reconstructed using 3, 6, and 12 numbers of perspectives and the ground truth image as captured by laying out the adhesive tape on a flat wall as shown in Figure 17(d).

To obtain a human-perspective evaluation of the quality of reconstruction, we asked 5 members of our lab to rate the distortion quality of the reconstructed images as compared to the ground truth image which was marked as 10/10. As can be noticed from the figures, the results show a clear increasing trend with higher number of perspectives available. This convinced us of our distortion analysis results and also convinced us that increasing number of perspectives via more cameras or more sensing cycles does reduce the image distortion experienced.

4.4. Optimal Sensor Selection. One interesting deliverable of our view based requirements analysis is the *optimal number* of sensors required for undertaking such a sensing task. Given the maximum acceptable distortion and the maximum sensing delay (i.e., number of sensing cycles allowed per combination), we can compute the minimum number of sensors which can fulfill the requirements for various views. Such an analysis is possible as there exist trade-offs between the image distortion and the number of perspectives from which the image is captured (as discussed in Section 3.2.2 and verified in previous experiment). These numbers of perspectives are in turn based on number of cameras being

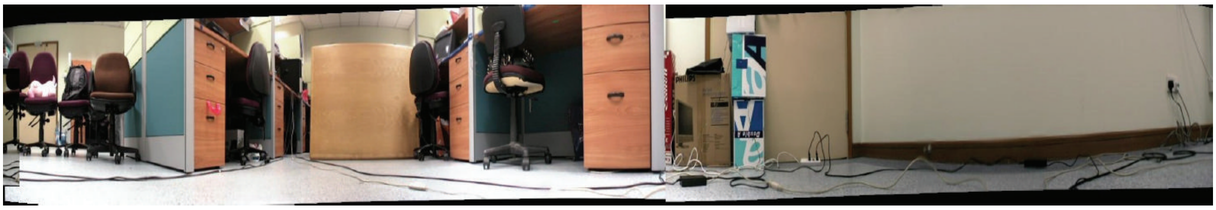


Figure 13: Environment mosaic.

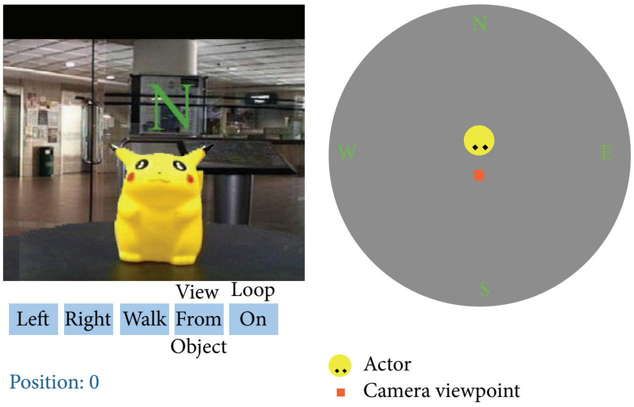


Figure 14: Snapshot of independent view preview.

employed and the number of sensing cycles allowed for one reconstruction. Thus, these three factors can be traded off as shown in Figure 18.

For example, if we are allowed a maximum distortion of 0.2 and a maximum delay of 2 sensing cycles, we can use Figure 18 to find the minimum number of cameras required as 4. Please note that this analysis can be used to find the least-costing combination *before* buying the physical equipment and thus helps in cutting down the project equipment cost.

4.5. Discussion. Based on our analysis and the described experimental results, we conclude the following:

(1) Decoupled sensing *can* actually capture 360° images of the actor and the environment.

(2) There exist feasibility regions ("FR"s) around the actor, where the cameras placed are significantly better suited for 360° image capture as opposed to other positions.

(3) These FRs also help in reducing the dynamic reconfiguration costs around a moving actor.

(4) An OAP-FR approach helps in further reducing this dynamic camera reconfiguration cost.

(5) Progressive realignment approach and delay allowance can help in increasing captured image quality by increasing the number of perspectives of image capture.

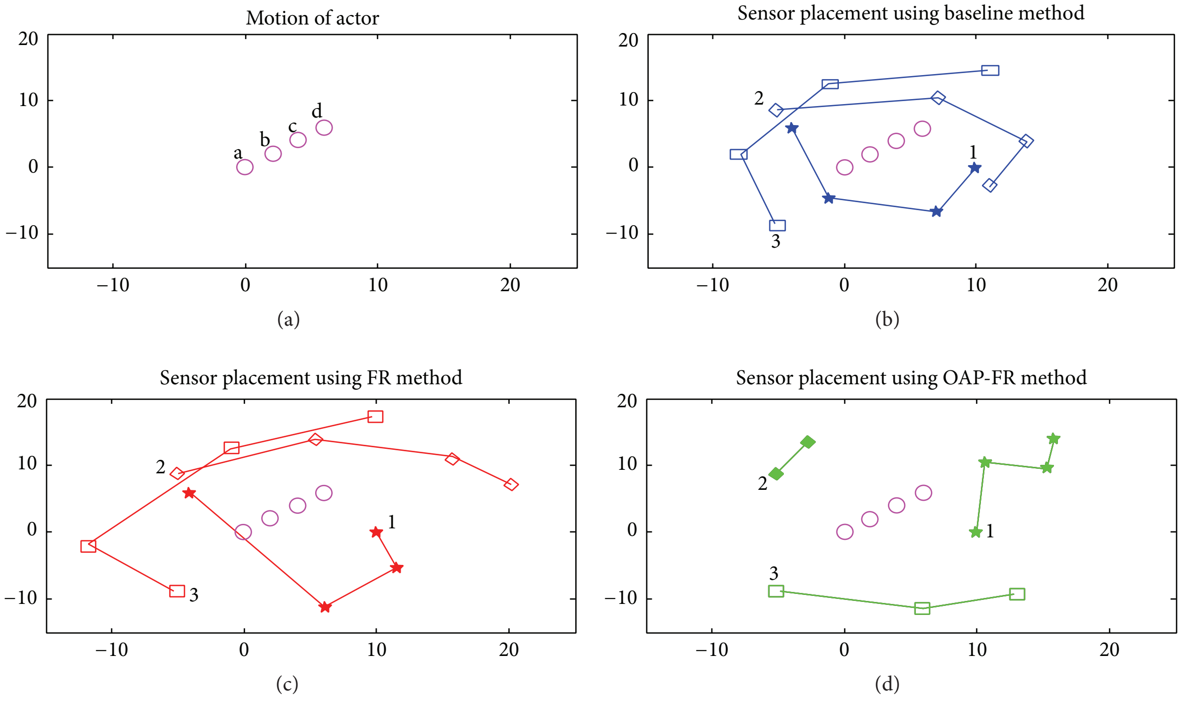
(6) There exist trade-offs between number of cameras, delay-allowance, and maximum acceptable distortion and this relationship can be used to find optimal sensor number.

While the proposed decoupled sensing paradigm is generic, we do realize that our above-mentioned experimental results were obtained under the assumptions like actor and environment convexity, single actor, no occlusions, limited depth, and nonlive data. Furthermore, our current experimental validation has focused on indoor small to medium scale environments. We acknowledge the complexities associated with large-scale scene capture and sensor movement in irregular environments (e.g., inaccurate movement of mobile camera sensors, avoiding obstacles, and lighting variation) but leave them outside the scope of the current work. However, in future work, we intend to progressively relax these assumptions for a more generic sensing application. Also, while we currently focused only on visual cameras, in the future we may employ IR cameras or other ranging sensors (for say a reconnaissance mission) to build an environment model and still find out what the actor shall observe if placed in such an environment.

We believe that this decoupled sensing approach (in its current or enhanced form) shall be very useful in many areas like reconnaissance and surveillance, telepresence, interactive television, military tasks, and analysis of sports and dance/performance videos. It can also be applied in pedagogy for placing students in the trainer's shoes and letting them experience the scene around them.

5. Conclusions

In this paper, we have analyzed a novel decoupled sensing paradigm which allows flexible sensing and reconstruction of any scene. This decoupled sensing allows viewers to choose to see the actor and the environment from any position and any



(a)

(b)

(c)

(d)

FIGURE 15: Comparison between trajectories for three sensor coordination strategies.

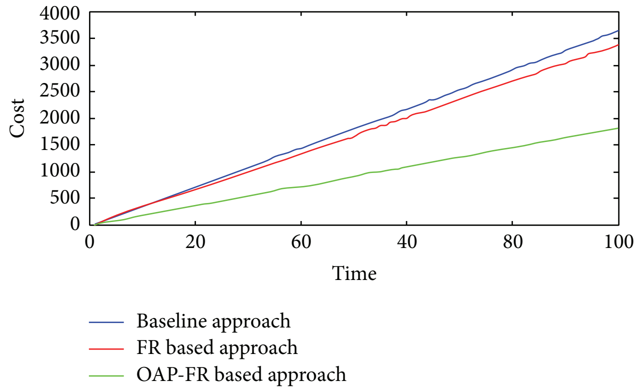
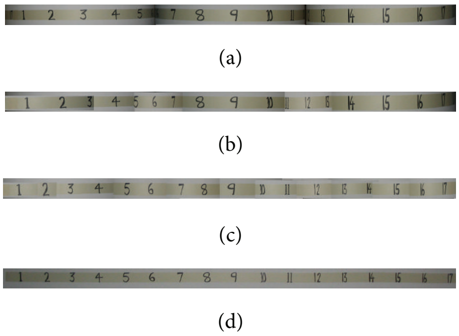


FIGURE 16: Comparison between costs for three sensor coordination strategies.



(a)

(b)

(c)

(d)

Set	Number of perspectives	Average score (max.: 10)
(a)	3	6.0
(b)	6	7.2
(c)	12	9.0
(d)	Ground truth	10.0

FIGURE 17: Effect of multiple perspectives on quality of images.

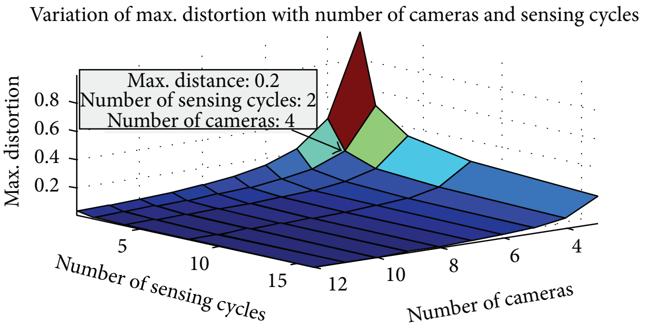


Figure 18: Trade-off.

direction. In order to undertake such sensing, the cameras need to be placed in certain positions (feasibility regions) around the actor which must change dynamically as the actor moves. We have provided mathematical analyses as well as practical results for the calculation of such feasibility regions and their benefits on sensing quality as well as on sensing cost. Lastly, we have also described a method for obtaining the minimum number of sensors required for undertaking such sensing tasks.

In our future work, we plan to handle "live" view generation for multiple, nonconvex actors in occluded environments. We also intend to study the use of multiple sensing modalities to increase the quality of the generated views while trying to decrease the costs incurred.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] T. Haenselmann, M. Busse, S. Kopf, T. King, and W. Effelsberg, "Multi perspective panoramic imaging," *Image and Vision Computing*, vol. 27, no. 4, pp. 391–401, 2009.

[2] G. Petrovic, D. Farin, and P. H. N. de With, "Toward 3D-IPTV: design and implementation of a stereoscopic and multiple-perspective video streaming system," in *Stereoscopic Displays and Applications XIX*, vol. 6803 of *Proceedings of SPIE*, International Society for Optics and Photonics, San Jose, Calif, USA, January 2008.

[3] M. Sharma, S. Chaudhury, and B. Lall, "Space-time parameterized variety manifolds: a novel approach for arbitrary multi-perspective 3D view generation," in *Proceedings of the International Conference on 3D Vision (3DV '13)*, pp. 358–365, IEEE, Seattle, Wash, USA, July 2013.

[4] August 2015, https://www.google.com/get/cardboard/jump/.

[5] August 2015, https://www.google.com/get/cardboard/apps/.

[6] P. H. Kelly, A. Katkere, D. Y. Kuramura, S. Moezzi, S. Chatterjee, and R. Jain, "An architecture for multiple perspective interactive video," in *Proceedings of the ACM International Conference on Multimedia*, San Francisco, Calif, USA, November 1995.

[7] T. Kanade, "The eye vision at super bowl and the virtualized reality system: 4d digitization of a time-varying real event and its application," in *Proceedings of the International Conference on Augmented, Virtual Environments and 3D Imaging*, Mykonos, Greece, June 2001.

[8] E. Borovikov, A. Sussman, and L. Davis, "A High Performance Multi-Perspective Vision Studio," in *Proceedings ACM International Conference on Supercomputing*, pp. 348–357, San Francisco, Calif, USA, June 2003.

[9] S. M. Seitz and J. Kim, "Multiperspective imaging," *IEEE Computer Graphics and Applications*, vol. 23, no. 6, pp. 16–19, 2003.

[10] https://360video.fb.com/.

[11] C. Zhang and T. Chen, "Active rearranged capturing of image-based rendering scenes—theory and practice," *IEEE Transactions on Multimedia*, vol. 9, no. 3, pp. 520–531, 2007.

[12] L. E. Parker, "Cooperative robotics for multi-target observation," *Intelligent Automation and Soft Computing*, vol. 5, no. 1, pp. 5–19, 1999.

[13] B. P. Gerkey and M. J. Matarić, "Sold!: auction methods for multirobot coordination," *IEEE Transactions on Robotics and Automation*, vol. 18, no. 5, pp. 758–768, 2002.

[14] F. Yang, M. Paindavoine, H. Abdi, and A. Monopoli, "Development of a fast panoramic face mosaicking and recognition system," *Optical Engineering*, vol. 44, no. 8, Article ID 087005, 2005.

[15] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, no. 1, pp. 83–97, 1955.

[16] August 2015, http://decoupledsensing.googlepages.com/.

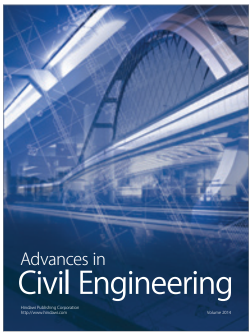
Journal of
Engineering

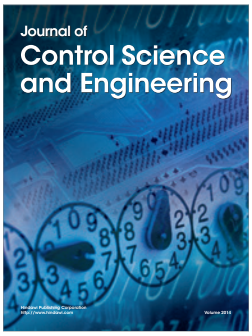
The Scientific
World Journal

International Journal of
Rotating
Machinery

Journal of
Sensors

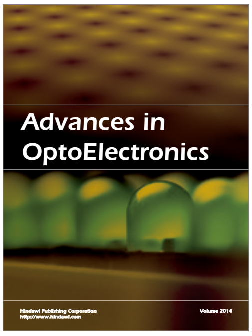
International Journal of
Distributed
Sensor Networks


Advances in
Civil Engineering


Journal of
Control Science
and Engineering

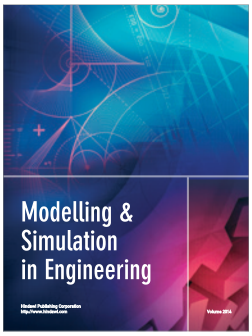
Journal of
Robotics

Journal of
Electrical and Computer
Engineering

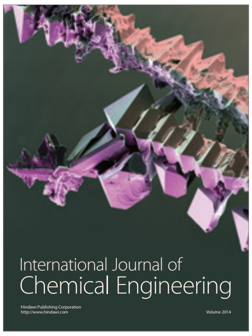

Advances in
OptoElectronics

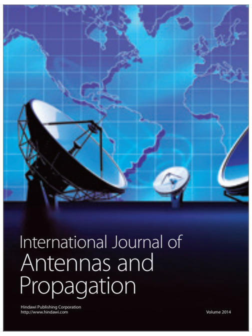
VLSI Design

International Journal of
Navigation and
Observation

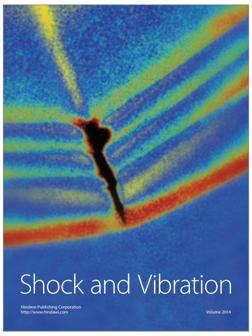

Modelling &
Simulation
in Engineering

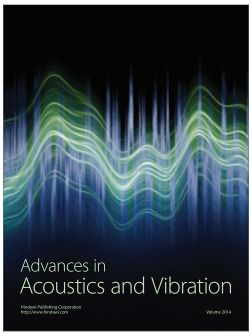
International Journal of
Aerospace
Engineering


International Journal of
Chemical Engineering


International Journal of
Antennas and
Propagation

Active and Passive
Electronic Components


Shock and Vibration


Advances in
Acoustics and Vibration