Temporal Fusion in Dynamic Linguistic Descriptions

Ozy Sjahputera, James M. Keller, *Fellow, IEEE*, Pascal Matsakis, Rajkumar Bondugula

***Abstract*—Previously, we introduced a method for estimating the linguistic direction of motion of an object in a scene by processing a temporal sequence of linguistic descriptions of its position relative to a stationary object. This method relies on the observation of "regular" descriptions generated through a fuzzy logic system based on the histogram of forces. These descriptions may not exist if the distance between the two objects is too close. In this paper, we use fuzzy sets to generate degrees of activation for the estimated directions of motion. Temporal fusion is applied on these values over a fixed time window using weights calculated from linguistic information on the distance between the objects. The new dynamic linguistic descriptions are determined based on the present as well as the past estimates.**

***Index Terms*—spatial relations, linguistic description of motion, scene interpretation, temporal fusion.**

I. Introduction

SPATIAL relations play an important part in scene interpretation. Various methods to model spatial relations between image objects have been proposed. For instance, the *histogram of angles* method was used in various studies [1][2][3]. In [4], Matsakis and Wendling introduced the concept of *the histograms of forces* (F-histogram), which generalizes and supersedes the angle histogram.

Linguistic descriptions of relative position is another important part in the study of scene interpretation. Winston [5] and Freeman [6] used linguistic terms to represent various primitive spatial relations between two objects. Freeman suggested the use of fuzzy sets to model these relations. Keller and Wang [7] introduced a fuzzy rule-base approach for linguistic scene description, where the relative position of two objects was represented using the angle histogram. However, the linguistic terms used were coarse and the approach failed to satisfy the *semantic inverse principle* in [6]. We introduced another fuzzy rule-base approach to linguistic scene description [8] based on the concept of F-histograms [4]. This approach offered a richer language that could be tailored to the user's needs. It also satisfied the semantic inverse principle.

Detecting and tracking moving objects in a scene has been widely discussed. The scene observation can be done via satellite images, aerial photographs, video sequences, GPS data, and telemetry observations. It provides a temporal sequence of static spatial information. This information may be extracted from each observation frame in the form of point coordinates, lines, or regions. In [9], objects were represented using minimum bounding rectangles. The center coordinates of the rectangles represented the object spatial information. Motion vectors were calculated from consecutive coordinates. In [10], spatial information was represented by linguistic descriptions. These *static linguistic descriptions* were generated using the method in [8]. The object's direction of motion was estimated by processing only the temporal sequence of the descriptions and an appropriate *dynamic linguistic description* of the motion was generated.

The static descriptions generated in [8] and used in [10] relied on the sole primitive directional relationships (RIGHT, ABOVE, LEFT, and BELOW). Moreover, our method in [10] assumed the observation of *regular descriptions*, which may not appear when the objects are too close and objects cannot be assimilated to points (ambiguous configurations). In this case, the system often gave erroneous estimates. In this paper, we introduce linguistic distance information as part of the linguistic descriptions used as inputs. We also assign a degree of activation to any dynamic linguistic description that can reasonably be used to represent the motion. Each description is represented by a fuzzy set and viewed as a hypothesis. The degree of activation generated at time t serves as the evidence in support of the hypothesis, hence allowing more than one hypothesis to be active. Keeping multiple hypotheses alive prior to making the final decision is in accordance with the *Principle of Least Commitment* proposed by Marr [11]. Each time the direction of motion is estimated, the linguistic distance information is used to calculate a weight reflecting our confidence of the estimate. Next, we introduce a temporal fusion mechanism to increase the accuracy of the estimation, especially when the distance between the objects gets too close. The mechanism fuses the evidence supporting current and past hypotheses within a fixed temporal window to obtain the correct direction of motion. The distance-based weight allows hypotheses taken when the two objects are farther apart to be more influential during the fusion process to compensate for possible erroneous hypotheses generated when the two objects pass each other at a very close distance. The hypothesis with the maximum evidence is selected to be the optimum description of motion. As in [10], we still maintain the following assumptions: one of the two objects is stationary, both objects can be assimilated to points and are relatively

The Office of Naval Research (ONR) grant N00014-96-0439 partially supported this work

O. Sjahputera (osb1e@mizzou.edu), J.M. Keller and R. Bondugula are with the Department of Computer Engineering and Computer Science, University of Missouri-Columbia, Columbia, MO 65211 USA

P. Matsakis was with the Department of Computer Engineering and Computer Science, University of Missouri-Columbia, Columbia, MO 65211 USA. He is now with the Department of Computing and Information Science, University of Guelph, Guelph, ON N1G 2W1, Canada

small in size, the motion follows a straight path with constant velocity, scene observation is taken over a regular time interval, and the observer's point of view is constant.

II. Background

A. Static Linguistic Descriptions

The input to our system is a temporal sequence $S=\{s_0,s_1,\cdots,s_t,\cdots,s_T\}$ where s_t is the *static linguistic description* observed at time t. The term *static* is used because the linguistic description represents the relative position of the *argument* object (A) with respect to the stationary *referent* object (B) without hinting any sense of motion among them. Under ideal configuration (objects are far enough apart and can be assimilated to points), there are 24 possible descriptions; each represents a conic region r centered on object B. These descriptions are called the *regular static linguistic descriptions* (L_r). Samples of the conic regions and their regular descriptions are illustrated in Fig. 1. The *primary* and *secondary* direction, δ_1 and δ_2, are the two primitive directions in {RIGHT, ABOVE, LEFT, BELOW} that best describe the position of A from B [8].

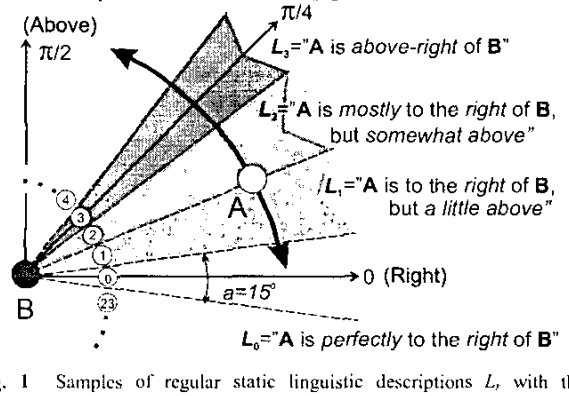

Fig. 1 Samples of regular static linguistic descriptions L_r with their corresponding regions r. Here, δ_1=RIGHT and δ_2=ABOVE.

B. Estimating the Direction of Motion

We record the number of observations for each L_r in S as $D_r=|\{t \mid s_t=L_r, s_t\in S\}|$. D_r and $D_{r'}$ are illustrated in Fig. 2. The direction of motion γ is estimated as

$$\gamma = \tan^{-1}[(\frac{N-1}{N+1})\cot(a)] + (r+\frac{1}{2})a, \qquad (1)$$

where a=15°, and currently $N=D_r/D_{r'}$ if r is "closer" to δ_1 ($\oplus$ transition), $N = D_{r'}/D_r$ otherwise ($\ominus$ transition). When a contiguous sequence s_t consists of non-regular descriptions with duration $\tilde{D}_r$, $N=(D_r+0.5\tilde{D}_r)/(D_{r'}+0.5\tilde{D}_r)$ is used for $\oplus$ transition, and $N=(D_{r'}+0.5\tilde{D}_r)/(D_r+0.5\tilde{D}_r)$ for $\ominus$ transition. The direction γ belongs to the interval $[-\pi/2, \pi/2]$ and is defined to be modulo π. For example, an angle of $3\pi/4$ corresponds to $\gamma=-\pi/4$. When this happens, the type of transition must be toggled to its opposite type to maintain directional consistency. When $\gamma=0$, the direction is perpendicular to δ_1 and parallel with δ_2 as shown in Fig. 2.

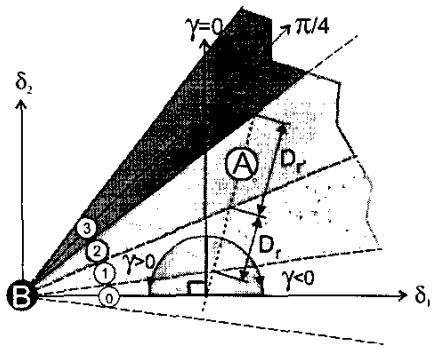

Fig. 2 Object A moves from region ⓪ to ③. The direction γ can be calculated as soon as A crosses from region ② to ③ using D_r and $D_{r'}$.

C. Dynamic Linguistic Descriptions

The *primary* and *secondary* motion direction, γ_1 and γ_2, are two primitive directions of motion that best describe the direction of motion of object A. They are determined from γ, δ_1 and δ_2 [10] as shown in Table I. The terms RIGHT, UPWARD, LEFT, and DOWNWARD are used for γ_1 and γ_2. They are comparable to the linguistic terms used for δ_1 and δ_2. From the example in Fig. 2 ($\gamma=-\pi/6$, $\oplus$ transition, δ_1=RIGHT and δ_2=ABOVE), Table I gives $\gamma_1=\delta_2$=ABOVE (UPWARD) and $\gamma_2=\delta_1$=RIGHT.

Table I Deriving γ_1 and γ_2, from γ, δ_1 and δ_2

γ	$\oplus$ transition		$\ominus$ transition	
	γ_1	γ_2	γ_1	γ_2
$\pi/4>\gamma\geq 0$	δ_2	$\delta_1+\pi$	$\delta_2+\pi$	δ_1
$\pi/2\geq\gamma\geq\pi/4$	$\delta_1+\pi$	δ_2	δ_1	$\delta_2+\pi$
$0>\gamma>-\pi/4$	δ_2	δ_1	$\delta_2+\pi$	$\delta_1+\pi$
$-\pi/4\geq\gamma>-\pi/2$	δ_1	δ_2	$\delta_1+\pi$	$\delta_2+\pi$

Let $\phi=\min(|\gamma_1|,\pi/2-|\gamma_1|)$ be the angle of motion viewed from γ_1 and γ_2 perspective. Based on ϕ, appropriate linguistic hedges are selected for γ_1 and γ_2 as shown in Table II. For compound directions, we use the term *"diagonally"* followed by both γ_1 and γ_2.

Table II Linguistic hedges for γ_1 and γ_2, ($a=\pi/6$)

ϕ	γ_1	γ_2
$a/2\geq\phi\geq 0$	*No hedge*	*No secondary Description*
$3a/2\geq\phi>a/2$	"Primarily"	"A little"
$5a/2\geq\phi>3a/2$	"Mostly"	"Somewhat"
$\pi/4\geq\phi>5a/2$	*Compound Direction*	

Any *dynamic linguistic description* (*DL*) is generated following the format "`A is moving [hedge of` γ_1`]` γ_1`, and [hedge of` γ_2`]` γ_2." The example in Fig. 2 yields $\phi=\pi/6$, hence *DL* for A is "`A is moving primarily upward, and a little to the right`". There are 24 possible dynamic descriptions DL_d, $d\in\{0,..,23\}$, where d represents a range of directions defined as a cone-shaped region shown in Fig. 3. Since DL_d captures the motion of object A, therefore the 24 cone regions are centered on A. See [10] for more discussion on γ and dynamic

descriptions.

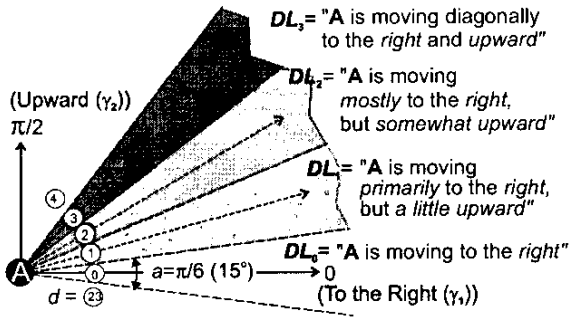

Fig. 3 Samples of dynamic linguistic descriptions and their regions (d). For region d=0-3, γ_1="To The RIGHT" and γ_2="UPWARD".

D. γ Normalization

In [10], the frame of reference for γ is based on the value of δ_1 and δ_2. With different values of δ_1 and δ_2, the same value of γ may actually point to two different directions in the observer's frame of reference. In this paper, we normalize γ to the observer's frame of reference. The normalized γ is denoted by γ^*: when $\gamma^*=0$, object A is heading to the right; γ^* belongs to $[0,2\pi]$ and increases in counter-clockwise direction. The normalization scheme shown in Table III assumes that object A follows the $\oplus$ transition. For $\ominus$ transition, apply $\gamma^*=\gamma^*+\pi$.

Table III γ normalization scheme

		δ_2			
		RIGHT	ABOVE	LEFT	BELOW
δ_1	RIGHT	N/A	$\gamma^*=\gamma+(\pi/2)$	N/A	$\gamma^*=(3\pi/2)-\gamma$
	ABOVE	$\gamma^*=-\gamma$	N/A	$\gamma^*=\gamma+\pi$	N/A
	LEFT	N/A	$\gamma^*=(\pi/2)-\gamma$	N/A	$\gamma^*=(3\pi/2)+\gamma$
	BELOW	$\gamma^*=\gamma$	N/A	$\gamma^*=\pi-\gamma$	N/A

III. Temporal Fusion of Dynamic Descriptions

A. Degree of Activation in Dynamic Descriptions

As the distance between objects A and B decreases, there is an increase in non-regular static descriptions. This increase causes the γ^* estimation error to increase as well, since the method in [10] was based on regular descriptions only. Furthermore, the direction γ^* was mapped onto a unique dynamic description, which limits the system flexibility in dealing with errors. Therefore, we now assign a *degree of activation* to any dynamic description that can reasonably be used to describe the motion of object A in direction γ^*. Each dynamic description is viewed as a hypothesis represented by a membership function μ_d. For this experiment, we use a triangular function with the peak located at the midpoint of the range d and 50% overlap with each adjacent function. Illustration of the μ_d is given in Fig. 4. From this example, we can see how we can have two hypotheses for γ^*, each with a degree of activation of $\mu_0(\gamma^*)$ and $\mu_1(\gamma^*)$.

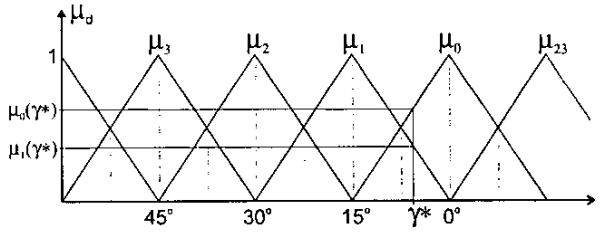

Fig. 4 The membership functions for dynamic description regions, μ_d. The estimated direction, γ^*, can reasonably be expressed using $DL_{d=0}$ (`A is moving to the right`) and $DL_{d=1}$ (`A is moving primarily to the right, but a little upward`), with degrees of activation of $\mu_0(\gamma^*)$ and $\mu_1(\gamma^*)$ respectively. Note $\mu_0(\gamma^*)>\mu_1(\gamma^*)$.

B. Distance Linguistic Information

We now add the distance information into the static linguistic descriptions we use as inputs. Six terms are employed: VERY_CLOSE ($Dist_0$), CLOSE ($Dist_1$), SOMEWHAT_CLOSE ($Dist_2$), SOMEWHAT_FAR ($Dist_3$), FAR ($Dist_4$), and VERY_FAR ($Dist_5$). Note that for our first approach, the linguistic distance information is crisply defined (of course, this need not be the case) by computing the minimum distance between any pair of points in A and B; the mapping of the minimum distance onto the corresponding linguistic term is done following the scheme in Fig. 5. The static description for the position of A relative to B is "`A is mostly to the right of B, but somewhat above. A is somewhat far from B.`"

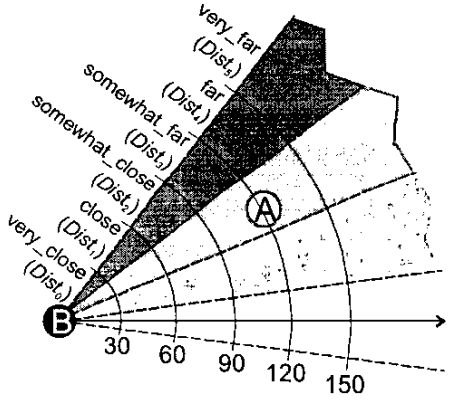

Fig. 5 Distance linguistic description: "`A is somewhat far from B`"

C. Processing the Distance Linguistic Information

As we noted previously, non-regular static linguistic descriptions tend to appear more often as the distance between the objects closes. The more non-regular description appear in S, the higher the difference between γ^* (estimated direction) and the actual direction of motion. Hence, the accuracy of the estimation is an increasing function of the distance between the two objects, which motivates us to introduce the linguistic distance information into the estimation of the direction of motion. First, we need a method to assign a value that represents the "overall" distance of object A from object B when γ^* is calculated. In Fig. 6, A moves from A_1 to A_2 by crossing regions ① and ②. At location A_2, as A crosses over

from ② to ③, γ^* is calculated using $D_r=D_2$ and $D_r=D_1$. During the motion from A_1 to A_2, the distance between object A and B are represented by linguistic distance descriptions $Dist_3$, $Dist_4$ and $Dist_5$, with total number of occurrences of $Obs(Dist_3)$, $Obs(Dist_4)$ and $Obs(Dist_5)$ respectively (and $Obs(Dist_0) = Obs(Dist_1) = Obs(Dist_2) = 0$).

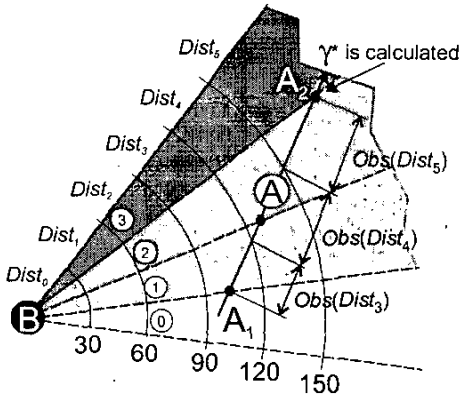

Fig. 6 Observation of distance information prior to calculating γ_t^*.

To calculate the overall distance Δ between objects A and B during the motion from A_1 to A_2, we model each $Dist_j$ using a fuzzy set μ_{Dist_j} as shown in Fig. 7. We use triangular membership functions that peak at the mid-point of each range defining $Dist_j$, and with 30% overlap with each adjacent function. The activation level of μ_{Dist_j} is:

$$\alpha_j = \frac{Obs(Dist_j)}{\sum_i Obs(Dist_i)} \tag{2}$$

The activation level α_j is then applied to μ_{Dist_j}. In this experiment, we use the *correlation-product* inference rule [12] and the resulting fuzzy sets are added to give us the aggregate fuzzy set. Δ, the overall distance associated with γ^*, is obtained by defuzzifying the aggregate fuzzy set. Here we use the *centroid defuzzification* scheme.

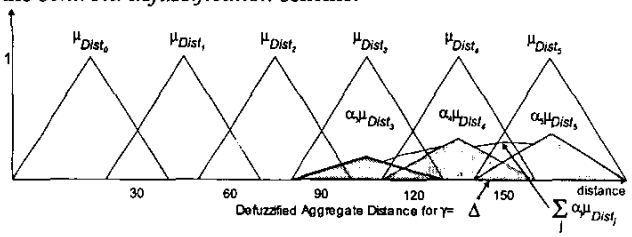

Fig. 7 The fuzzy sets for $Dist_j$ and how Δ for example in Fig. 6 is calculated.

D. Temporal Fusion

As explained in [10], at time t, the estimated direction of motion $\gamma^* = \gamma_t^*$ is calculated using only the observation information from the last two conic regions r and r'. Information from other regions prior to r or past estimates γ_{t-k}^* for $k \geq 1$ were not taken into account during the calculation for γ_t^*. To take advantage of the past information and the knowledge we have on the correlation between the estimation accuracy and the distance between objects, we introduce a temporal fusion function $H(t,k,d)$ that operates over a temporal window k on hypothesis μ_d at time t.

$$H(t,k,d) = \sum_{i=0}^{k-1} w_{t-i}\mu_d(\gamma_{t-i}) \tag{3}$$

For our experiments, we use $k = 5$ as the width of our temporal window. The weight $w_t \in [0,1]$ represents our confidence on the accuracy of γ_t^*. It is an increasing function of the overall distance Δ_t between objects A and B throughout the observation period that leads to the calculation of γ_t^*. Here, w_t is calculated as $w_t=\Delta_t/165$, where 165 is the midpoint of the support for μ_{Dist_5}.

As discussed in section III.A, $\mu_d(\gamma_t^*)$ is the degree of activation for each hypothesis, d is the sector number corresponding to a specific dynamic linguistic description (see Fig. 3) and μ_d is the membership function representing the dynamic description (see Fig. 4). Now, $H(t,k,d)$ is calculated for all $d \in \{0,..,23\}$. $H(t,k,d)$ can be viewed as the total weighted evidence that supports the hypothesis that the dynamic linguistic description d is the most appropriate description for the motion of object A. At time t, the linguistic expression FDL_t selected for describing the motion is the dynamic description d whose $H(t,k,d)$ is the highest.

$$FDL_d^t = \arg\max_d (H(t,k,d)) \tag{4}$$

IV. Results

Our simulation program is implemented using the C language with OpenGL graphic library and the Glut interface. Objects A and B are circular. The radius for each object is determined individually. The user sets the coordinates of object B and the starting and ending coordinates that define the path for object A. The path is traced using the Bresenham's line algorithm. First, object A passes object B at a "very close" distance. We expect to see inaccurate dynamic descriptions generated when A is at the closest position to B. The temporal fusion method is expected to overcome this problem and return the correct dynamic description, providing that adequate evidence to support the correct dynamic description is available prior to object A getting too close to B. In the following experiment, we move the path of A farther away from B to show that the accuracy of dynamic descriptions generated by a single estimation of γ_t^* increases as the distance between objects A and B increases.

A. Experiment 1: A is "very_close" to B

We now show examples of estimation errors that occur when the distance between objects A and B is very close and how the temporal fusion approach recovers the correct dynamic description (Fig. 8). A white line marks the path of the motion at 155°. The expected dynamic description is "A is moving *mostly* to the *left*, but *somewhat upward*" (d=10). The object radius is 10 pixels. Small

circles along the path identify the spots where transitions of regular static descriptions occur and dynamic descriptions are generated. The circles are called *transition points* and are numbered in the order of occurrence (1 to 11). These circles have smaller diameter than that of the actual object. Concentric rings centered on B mark the boundaries for distance linguistic descriptions.

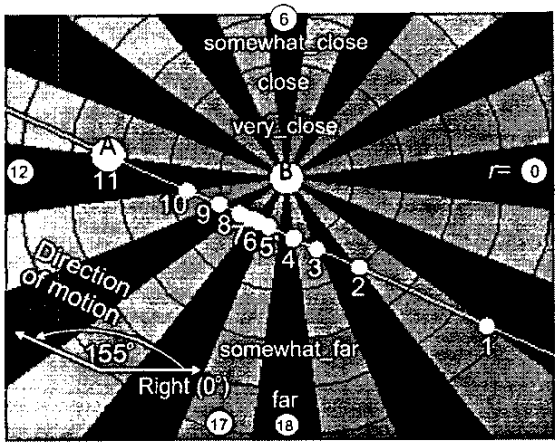

Fig. 8 Experiment 1: object A is heading to direction 155°. The worst estimation errors are observed when A is "very_close" to B (transition points 5 and 6). Results are shown in Table IV.

In Table IV, column t lists the order of transition points along the path of motion. Values t=1 and t=2 are not included, since we do not have sufficient observations to calculate γ_t* (for more information see [10]). Column DL^t_d **(1)** and DL^t_d **(2)** show the dynamic region number d of the first and second choice of dynamic descriptions generated from a single γ_t* estimation. Their degrees of activations are given in μ_d **(1)** and μ_d **(2)**. Column FDL^t_d **(1)** and FDL^t_d **(2)** lists the region number d for the first and second choice dynamic description generated using the temporal fusion mechanism. Their degrees of activations are given in **H (1)** and **H (2)**. Region d=11 represents "Object A is moving *primarily* to the *left*, but *slightly upward*", and d=10 represents "Object A is moving *mostly* to the *left*, but *somewhat upward*".

Table IV Results from experiment 1 in Fig. 8

t	γ_t*	DL^t_d (1)	μ_d (1)	DL^t_d (2)	μ_d (2)	FDL^t_d (1)	H (1)	FDL^t_d (2)	H (2)	w_t
3	154.2	10	0.719	11	0.281	10	0.349	11	0.137	0.486
4	161	11	0.735	10	0.265	10	0.405	11	0.290	0.208
5	197.9	13	0.807	14	0.193	10	0.405	11	0.290	0.097
6	121.3	8	0.914	9	0.086	10	0.405	11	0.290	0.091
7	143.5	10	0.569	9	0.431	10	0.456	11	0.290	0.091
8	156.5	10	0.569	11	0.431	11	0.192	10	0.159	0.091
9	154.3	10	0.716	11	0.284	10	0.168	13	0.078	0.091
10	161.8	11	0.786	10	0.214	10	0.199	11	0.176	0.142
11	155.5	10	0.634	11	0.366	10	0.395	11	0.289	0.309

The dynamic description generated from a single γ_t* may not capture the true direction of motion when A and B is too close. The worst error are found at t=5 and t=6 where A is "very_close" to B. The transition points along the path of motion (see Fig. 8) mark the location where the next regular static description is first observed. Ideally, this should occur right at the boundary between the two adjacent regions. However, at t=5, the transition is not detected until A is well into the region r=17. This indicates the presence of non-regular static descriptions during the transition from r=18 to r=17. The adverse effect of non-regular descriptions on the accuracy of γ_t* is lessened by distributing the evidence of non-regular description $\tilde{D}_r$ to both regular regions (r=18 and r=17) as discussed in section II.B. This method may not work well if number of non-regular descriptions $\tilde{D}_r$ gets too large relative to the number of regular descriptions observed. Between t=4 and t=5, we observe 6 non-regular descriptions and only 10 regular descriptions, hence the value for $D_{r=18}$ (see Fig. 2) is inaccurate, because it is larger than it should be (some portion of r=17 has been erroneously identified to belong to r=18). This causes the value of γ_t* at t=5 (using $D_{r'} = D_{r=18}$ and $D_r = D_{r=19}$) to be off by 42.9° (the largest in this experiment) compared to the actual direction of motion (155°). The inaccuracy of $D_{r=18}$ also propagates to γ_t* calculated at t=6 (now use $D_{r'} = D_{r=17}$ and $D_r = D_{r=18}$) which results in an error of 33.7°, the second largest estimation error in this experiment. Excluding t=5 and t=6, the average γ_t* error for the whole motion is just around 4°.

At t=5, the dynamic description generated from a local γ_t* is "Object A is moving *primarily* to the *left*, but *a little downward*" (d=13). It still correctly captures the value for γ_1 (to the left) but fails to capture the correct value for γ_2 (upward). The dynamic description t=6 is "A is moving *mostly upward*, but *somewhat* to the *left*" (d=8). It fails to express the correct direction of motion; it reverses the order of γ_1 and γ_2. On the other hand, almost all FDL^t_d obtained from the temporal fusion method return the correct dynamic description throughout the motion, thanks to the distance-based weight w_t used in the fusion process. The only error in FDL^t_d occurs at t=8: "Object A is moving *primarily* to the *left*, but *a little upward*" (d=11). However, from Table IV, we see that the next highest total evidence $H(t,k,d)$ corresponds to the correct dynamic description "A is moving *mostly* to the *left*, but *somewhat upward*" (d=10). Moreover, the two $H(t,k,d)$ values are relatively close: 0.192 for d=11, 0.159 for d=10. The direction 155° is in a region where $\mu_{d=10}$ and $\mu_{d=11}$ overlap, hence both $DL_{d=10}$ and $DL_{d=11}$ make good candidates for the description of this motion. We expect $DL_{d=10}$ to carry a higher degree of confidence because direction 155° is closer to the peak of $\mu_{d=10}$. However, it is not completely unreasonable to use $DL_{d=11}$ to describe this motion either. The true direction 155° is close to the point of maximum uncertainty between the two hypotheses (occurs at direction 157.5° where $\mu_{d=11}(\gamma_t^*) = \mu_{d=10}(\gamma_t^*) = 0.5$). Hence, we feel that the observed error from the temporal fusion approach is within an acceptable limit.

B. Experiment 2: A is "close" from B

Object A moves to heading 161.4°. It passes object B at a "close" distance (see Fig. 9). The expected dynamic description is "Object A is moving *primarily* to

the *left, but a little upward"* (d=11). From Table IV, we note that errors in DL^t_d(1) are observed at t=5, t=6, and t=7. However, these errors are less severe than those observed in experiment 1. At t=5, the dynamic description is "Object A is moving to the *left"* (d=12), while at t=6 and t=7 the description is "A is moving *mostly* to the *left, but somewhat upward"* (d=10). Both d=12 and d=10 are adjacent to the expected description of d=11. However, for this heading, only $\mu_{d=10}$ and $\mu_{d=11}$ should be activated, which makes d=10 to be the next best description after d=11. The maximum estimation error is just above 14°, much smaller compared to that of experiment 1. This indicates a very small number of non-regular descriptions observed at this distance. Visually, we can confirm this notion by noting that all transition points along the path of motion are identified just around the boundary between two adjacent regions r. All dynamic descriptions generated using the temporal fusion approach yield the correct values (d=11) as shown in column FDL^t_d(1). The temporal fusion results also confirms that the dynamic description d=10 is the second best description for this motion, as shown in column FDL^t_d(2).

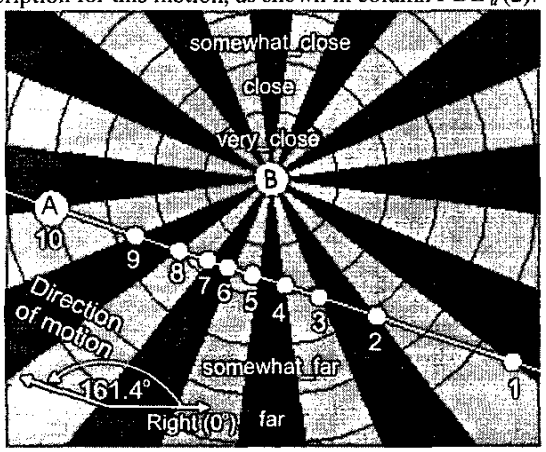

Fig. 9 Experiment 2: object A is heading to direction 161.4°. The closest distance between A and B is "close". Both the number and severity of errors in this experiment are less compared to those observed during Experiment 1.

Table V Results from experiment 2 in Fig. 9

t	γ_t^*	DL^t_d (1)	μ_t (1)	DL^t_d (2)	μ_t (2)	FDL^t_d (1)	H (1)	FDL^t_d (2)	H (2)	w_t
3	160.4	11	0.695	10	0.305	11	0.455	10	0.2	0.655
4	163.8	11	0.923	10	0.077	11	0.797	10	0.228	0.370
5	174.8	12	0.655	11	0.345	11	0.891	10	0.228	0.273
6	152.3	10	0.847	11	0.153	11	0.933	10	0.459	0.273
7	147.1	10	0.807	9	0.193	11	0.933	10	0.679	0.273
8	164.4	11	0.96	10	0.04	11	0.739	10	0.490	0.273
9	163	11	0.866	10	0.134	11	0.664	10	0.503	0.307
10	162.2	11	0.814	10	0.186	11	0.950	10	0.590	0.467

V. Conclusion and Future Work

In [10], we introduced a method for motion estimation and description using linguistic expressions as the only inputs. However, the method does not work well if the distance between the considered objects A and B is too close. With this in mind, we added linguistic distance information to the static linguistic descriptions used as inputs. Then, we employed a fuzzy set to represent each range of motion directions associated with each possible dynamic description. This allowed us to assign degrees of activation to multiple dynamic descriptions. We introduced a temporal fusion mechanism into the system, which took advantages of the distance information and the degrees of activation given to multiple candidates for dynamic descriptions. From our examples using synthetic data, we showed that the temporal fusion approach increased the system robustness in dealing with errors attributed to distance-induced uncertainties. The current system still relies on regular static descriptions to identify the transition from one cone-region to the next. With the help of distance information and temporal fusion, our current system is able to robustly handle some non-regular descriptions (linguistic descriptions of ambiguous configurations [8]). However, if the distance between the two objects remains too close during the motion, then it is possible that there is no regular static description generated at all. As the next step, we will extend the system's capability to deal with both regular and non-regular static descriptions to increase its reliability and reduce its sensitivity to objects' sizes, shapes and the distance between them. We also plan to add the ability to detect and describe changes in direction of motion.

References

[1] K. Miyajima and A. Ralescu, "Spatial Organization in 2D Segmented Images: Representation and Recognition of Primitive Spatial Relations", *Fuzzy Sets and Systems*, vol. 65, iss. 2/3, pp. 225-236, 1994.

[2] R. Krishnapuram, J. M. Keller and Y. Ma, "Quantitative Analysis of Properties and Spatial Relations of Fuzzy Image Regions", *IEEE Trans. on Fuzzy Systems*, vol. 1, no. 3, pp. 222-233, 1993.

[3] X. Wang and J. M. Keller, "Human-Based Spatial Relationship Generalization Through Neural/Fuzzy Approaches", *Fuzzy Sets and Systems*, vol. 101, no. 1, pp. 5-20, 1999.

[4] P. Matsakis, L. Wendling, "A new way to represent the relative position between areal objects," *IEEE Trans. PAMI*, vol. 21, pp. 634-643, 1999.

[5] P. Winston and "Learning Structural Descriptions from Examples", in P. Winston (ed.), *The Psychology of Computer Vision*, McGraw-Hill, New York, 1975.

[6] J. Freeman, "The Modeling of Spatial Relations", *Computer Graphics and Image Processing*, vol. 4, pp. 156-171, 1975.

[7] J. M. Keller and X. Wang, "A Fuzzy Rule-based Approach for Scene Description Involving Spatial Relationships", *Computer Vision and Image Understanding*, vol. 80, pp. 21-41, 2000.

[8] P. Matsakis, J.M. Keller, J. Marjamaa, O. Sjahputera, "Linguistic description of relative positions in images," *IEEE Trans. SMC-B*, vol. 31, no. 4, pp. 573-588, August 2001.

[9] K. Beard, H.M. Palancioglu, "Estimating Positions and Paths of Moving Objects", *Proc. 7th Int'l Workshop on Temporal Representation and Reasoning*, pp. 155 –162, 2000.

[10] O. Sjahputera, P. Matsakis, J.M. Keller, R. Bondugula, "Linguistic Descriptions for an Object in Motion," in *Proc. North American Fuzzy Info. Proc. Soc.* Annual *Conf.*, New Orleans, 2002, pp. 243-248.

[11] D. Marr, *Vision*. San Francisco, CA: Freeman and Company, 1982.

[12] B. Kosko, *Neural Networks and Fuzzy Systems*. Englewood Cliffs, New Jersey: Prentice Hall, 1992, pp. 312.