

大規模映像解析システム向けのサーバ割当手法

Server resource allocation and control method for large-scale video analysis systems

岩松洋介[†] 小山和也[†]

YOSUKE IWAMATSU[†] KAZUYA KOYAMA[†]

1. はじめに

大量の監視カメラ映像のリアルタイム解析等、映像解析技術を活用した大規模システムの需要が増大している。

映像解析は、処理の負荷が高く多数のサーバを必要とするが、前後の映像を見て解析を行うため、実行中のサーバ変更は通常できない。負荷の集中による応答性能劣化を避けつつ、必要なサーバ資源を抑えることが難しく、応答要求の充足と実行効率を両立するサーバ割当が求められる。

本論文では、負荷の集中を適切に扱うため、応答要求の異なる解析のサーバ内におけるバッファ優先度制御と、それを前提として解析をバランスよく配置するサーバ間の割当を組み合わせた手法を提案する。実際の都市監視システムの負荷データを元に評価し、応答要求の充足と実行効率の向上の効果を確認した。

2. 映像解析システム効率化の課題

映像解析システムでは、カメラ映像に対し、人物検出や顔照合等の解析を行い、解析結果をユーザに提供する。リアルタイムでのアラート表示、蓄積した解析結果の検索等、様々な応答性能が求められるという特徴がある。

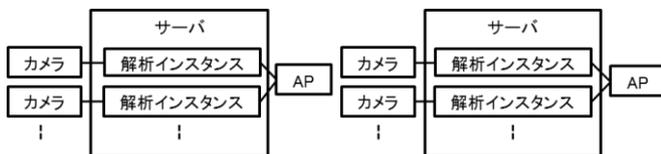


図1 映像解析システムの構成

映像解析システムの一般的な構成を図1に示す。解析処理の実行主体を解析インスタンスと呼ぶ。映像解析は、通常、前後のフレームの状態を参照するため、フレーム毎に処理を分散させることができず、一カメラのデータを一解析インスタンスで処理し、複数の解析インスタンスを複数のサーバに固定的に割り当てて運用される。

個々の解析インスタンスは、映っている人数等映像内容に応じて負荷が変動する。一サーバ内で負荷が集中すると、負荷溢れが発生、解析処理が遅延し、応答要求を満たせなくなるため、解析インスタンスのサーバ割当が問題となる。

通常、負荷溢れを回避するには、最大負荷割当が用いられる。この手法では、解析インスタンス毎のピーク時の最大負荷を予測し、負荷が集中しても耐えられる十分なサーバを割り当てる。しかし、必要なサーバ数が極めて大きい。

サーバ数を削減したい場合、ベストエフォートで負荷溢

れの発生を抑える手法が用いられる。解析インスタンス毎の平均負荷を予測し、サーバ間で平準化する、平均負荷割当が代表的である。しかし、負荷の変動を考慮しないため、予測が正しくとも、負荷溢れの可能性を排除できない。

従って、個々の解析インスタンスが応答要求を満たすことを保証しつつ、実行効率を向上させ、必要なサーバ資源を削減するサーバ割当の実現が課題である。

3. バッファ利用の解析制御・サーバ割当

本論文では、サーバ内でバッファを用いた優先度制御を行いつつ、サーバ間では、バッファ優先度制御の実施を前提に、応答性能を維持しつつ効率を高めるように割当を行う手法を提案する。これによって、負荷予測が正しい場合の応答要求の充足を保証しつつ、必要なサーバを削減する。

3.1 サーバ内バッファ優先度制御

サーバ内バッファ優先度制御の概要を図2に示す。応答要求の低い解析インスタンスに対し、前段にバッファを配置し実行タイミングを調整し、応答要求の高い解析インスタンスの空き時間を利用して実行するよう、入力データ量を調節する。

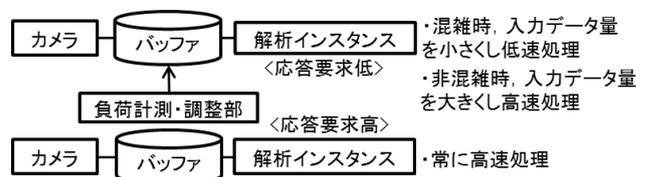


図2 サーバ内バッファ優先度制御

3.2 バッファ優先度制御を考慮したサーバ割当

バッファには、負荷を時間方向に平準化する効果がある(図3)。ある時間幅でバッファリングする場合、通常は負荷が変動するところを、その時間幅内では一定として扱える。時間幅を小さくすると、この値はピーク時の最大負荷に近づき、時間幅を大きくすると平均負荷に近づく。時間幅が大きいくほど、必要なサーバ処理能力を低減できる。

また、優先度制御には、応答要求の異なる解析の実行タイミングをずらす効果がある。瞬間的に負荷が集中した場合、応答要求の高い解析インスタンスのみが実行される。ある時間幅を考えると、応答要求が高く、許容するTATがその時間幅より短い解析インスタンスについてのみ、負荷の合計に対応できる処理能力を用意すればよい。時間幅を上下させ、各時間幅において必要な処理能力が用意できていれば、全体の応答要求を満たすことができる。

[†] 日本電気株式会社, NEC Corporation

つまり、短い時間幅では、考慮すべき個々の負荷は大きいものの、解析インスタンスの数は少ない。長い時間幅では、解析インスタンスの数は多いものの、個々の負荷は小さい。この性質を利用することで、最大負荷割当と同様に、負荷予測が正しければ負荷溢れを回避しつつ、必要なサーバ資源を削減できる。

本手法では、バッファによる負荷の時間方向平準化の効果と、優先度による実行タイミング制御の効果について、時間幅毎最大負荷、時間幅毎余剰能力の概念を導入し、一般的な形にモデル化する。そして、応答要求に応じて複数の時間幅を設定し、サーバ毎、時間幅毎の余剰能力を算出し、それらを均等にする割当パターンを選択する。以下、本手法の特徴となるモデルについて説明する。

時間幅毎最大負荷

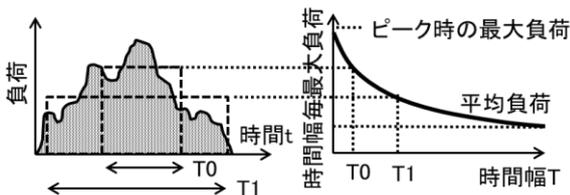


図 3 時間幅毎最大負荷

時間幅毎最大負荷 $Ln(T)$ は、解析インスタンス n の負荷の時系列変化が与えられた場合の、時間幅 T について、最混雑時間帯の負荷の単位時間平均である(図 3)。 $Ln(T)$ は、時間幅 T の増加に伴い減少し、以下の式で近似できる

$$Ln(T) = a - b \times \log T \quad (a, b \text{ は定数})$$

時間幅毎余剰能力

時間幅毎余剰能力 $Ci(T)$ は、時間幅 T におけるサーバの余剰能力を表す。サーバの処理能力を C 、解析インスタンス n の許容 TAT を Rn とする。

$$Ci(T) = 1 - \left\{ \sum_n Ln'(T) \right\} / C$$

但し、
 $Ln'(T) = Ln(T)$ [if $Rn \leq T$]
 $Ln'(T) = 0$ [if $Rn > T$]

4. 評価と考察

4.1 評価方法

提案手法にもとづき、バッファ優先度制御およびサーバ割当を実装した。解析制御の実装には、文献[1][2]に記載の解析制御ミドルウェア (ACM) を用いた。評価用データとして、ある都市監視の実証実験にて得られた、人物行動解析、顔特徴解析、物体異常検知等の映像解析の負荷データを用いた。一日の負荷データを予測に用い、サーバ割当に使用した。また、別の一日の負荷データを用いてテスト用の模擬負荷を生成し、TAT 充足率を計測した。

4.2 TAT 充足率の比較

表 2 は、表 1 の解析インスタンス 1~10 を、コア数 2 のサーバ 3 台 A~C で実行した場合についての、サーバ割

当結果および TAT 充足率の比較である。平均負荷割当では、応答要求の違いや負荷変動を考慮せずに割当を行うため、サーバ C に応答要求の高い解析が集中し、TAT を充足できなかった。提案方式では、応答要求の高い解析と低い解析がバランスよく割り当てられ、それぞれ TAT を充足した。

表 1 解析インスタンスの応答要求タイプ

応答要求タイプ	許容 TAT	解析インスタンス
低	3600(秒)	1, 2, 3
高	10(秒)	4, 5, 6, 7, 8, 9, 10

表 2 サーバ割当結果および TAT 充足率の比較

	サーバ割当結果			TAT 充足率	
	サーバ	解析インスタンス		低	高
		低	高		
平均負荷割当	A	2	5, 10	1.0	1.0
	B	1, 3	7	1.0	1.0
	C	-	4, 6, 8, 9	-	0.26
提案方式	A	3	4, 8, 9	1.0	1.0
	B	2	5, 10	1.0	1.0
	C	1	6, 7	1.0	1.0

4.3 必要サーバ資源の比較

図 4 は、解析インスタンス 1~10 について、応答要求の低い解析の数を変動させ、必要なサーバのコア数を算出した結果である。提案方式により、一部の解析インスタンスで遅延を許容するだけで必要コア数を削減できる。全て応答要求の高い場合が、最大負荷割当に該当する。それと比較し、最大で約 40%削減可能と考えられる。

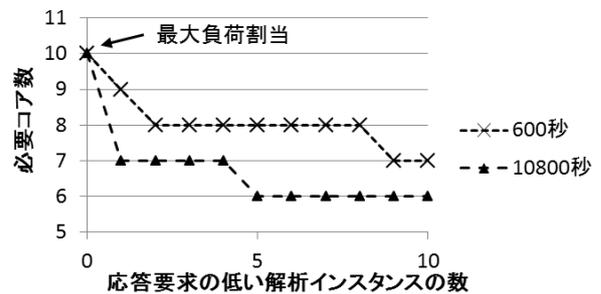


図 4 必要コア数の見積もり

5. まとめ

本論文では、映像解析システムの大規模化に向けたサーバ割当手法を提案した。実映像解析システムの負荷データを元に評価を行い、応答要求の充足と実行効率を両立し、必要なサーバ資源を削減する効果を確認した。

今後は、様々な実映像解析システムに対して本手法の適用可能性の検証を進め、大規模映像解析システムの拡大を推進していきたい。

参考文献

[1] 小山和也ほか, “大規模映像解析システム向けの解析制御ミドルウェアの提案”, 第 11 回情報科学技術フォーラム (2012)
 [2] Takeshi Arikuma et al., ”Analysis control middleware for large-scale video surveillance”, Advanced Video and Signal Based Surveillance (2013)